

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2003, 03:34:33 ; Search time 1115 Seconds
(without alignments)
4575.407 Million cell updates/sec

Title: US-09-955-807-1_COPY_59_373

Perfect score: 315
Sequence: 1 atgtctgtattcttcgagcc.....atgtcttcacgagccac 315

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40.2	12.8	746	12	BG024363 602274356
2	39	12.4	764	9	AL547921 AL547921
3	36.8	11.7	893	12	BG253488 602364213
4	35.8	11.4	709	13	BM015593 603641823
5	35.6	11.3	362	12	BF879524 PM4-ET015
6	35.6	11.3	389	13	BM030759 495261 MA

Result No.	Score	Query Match	Length	DB ID	Description
7	34.8	11.0	648	10	AM845712 MR0-CT006
8	34.8	11.0	1088	17	AL101449 Drosophila
9	34.6	11.0	1079	13	B1823340 603041251
10	34.4	10.9	518	9	A1800420 t114D10.x
11	34.2	10.9	624	17	A2981841 2M0262123
12	34.2	10.9	675	10	BE311470 601142205
13	34.2	10.9	1300	14	BO931641 AGENCOURT
14	34.2	10.8	351	12	BE929882 RC5-GN004
15	34	10.8	413	9	AA740754 n203c06.s
16	34	10.8	599	17	AQ495774 HS-5219_B
17	33.8	10.7	450	9	AA769321 n239g07.s
18	33.8	10.7	479	12	BG481055 602529036
19	33.8	10.7	529	12	BG788930 SEAMC008
20	33.6	10.7	1341	17	AG093150 Pan t1rog1
21	33.4	10.6	437	9	A1051406 ox40h11.s
22	33.4	10.6	510	17	AQ077664 CIT-HSP-2
23	33.4	10.6	606	17	BH038539 RPI-24-3
24	33.4	10.6	679	10	BE273345 601142647
25	33.2	10.5	295	10	AM58199 42179 MAR
26	33	10.5	379	9	AA485754 ab10802.r
27	33	10.5	424	9	AA807714 n331f07.s
28	33	10.5	598	10	AV720639 AV720639
29	33	10.5	640	13	B1046489 MR3-FN020
30	33	10.5	646	10	AV697956 AV697956
31	33	10.5	672	17	A2343652 1M0077F13
32	33	10.5	754	14	BO159046 NF106E06P
33	33	10.5	1005	17	CNS020F2
34	32.8	10.4	690	12	BG108181 602280048
35	32.8	10.4	827	12	BF165435 601777326
36	32.8	10.4	877	12	BG542440 602569515
37	32.6	10.3	338	13	BM193389 TCBAPE198
38	32.6	10.3	346	12	BF353335 QV1-HT063
39	32.6	10.3	378	10	AM502891 UT-HF-BNO
40	32.6	10.3	446	14	W42502 zc84d04.s1
41	32.6	10.3	560	17	A2288367 RPI-23-4
42	32.6	10.3	563	14	N90521 zbd1b08.s1
43	32.6	10.3	824	12	BE903653 601675142
44	32.6	10.3	976	10	BB622030 BB622030
45	32.6	10.3	1154	14	BQ066068 AGENCOURT

ALIGNMENTS

RESULT 1
LOCUS BG024363 746 bp mRNA linear EST 24-JUN-2001
DEFINITION 602274356F1 NIH_MGC_85 Homo sapiens CDNA clone IMAGE:4362493 5',
mRNA sequence.
ACCESSION BG024363
VERSION BG024363.1 GI:12409862
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 746)
NIH-MGC http://mgi.nci.nih.gov/
TITLE Unpublished (1999)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
CONTACT: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Louis Staudt, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL).
DNA Sequencing by: Incyte Genomics, Inc.
DNA distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10006 row: p column: 14
High quality sequence start: 2
High quality sequence stop: 540.
Location/Qualifiers

Db 262 CTGGCCCTCTCTCCACATGTATCATGAGCGTGGCCCTGTCTACACACCTTCGGAAACCTGGG 203

Qy 100 GCAGGAGTATATAGAGGCCACATCCCAAGATGAGCCTGGAAACAATCTCTCTCTGCCCCCTCA 159

Db 202 CGTAGAGGAATGACAGGCTCTCTCCCTGCCCCCTGCACTGCCCCCTGTGCTCTGCTCTGGA 143

Qy 160 CCAGCAGTACAGAT 175

Db 142 AGGCCGTGGGTCCGGTT 127

RESULT 4	LOCUS	DEFINITION
BM015593	709 bp	EST 30-OCT-2001
BM015593	709 bp	linear
603641822F1	NIH_MGC_87	Homo sapiens cDNA clone IMAGE:5418029 5',

ACCESSION	BM015593
VERSION	BM015593.1
	GI:16529947

SOURCE ORGANISM

REFERENCE	Eukaryote: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (Pages 1 to 709)
TITLE	NIH-MGC http://mgc.ncl.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DRP
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
<http://image.lnl.gov>
 Plate: LLAM12066 row: m column: 06
 High quality sequence stop: 154.

FEATURES
 Location/Qualifiers
 I..709

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5418029"
/clone_1ib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies
Note: this is a NIH_MGC library."

```

[illegible]

RESULT 5	
BF879524/c	
LOCUS	BF879524 362 bp mRNA linear EST 17-JAN-2001
DEFINITION	P44-ET0154-251100-005-d09 ET0154 Homo sapiens CDNA, mRNA sequence.
ACCESSION	BF879524
VERSION	BF879524.1 GI:12269654

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 362) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Bionesi,M.R., Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsunuma,A., Bala,G.S., Simpson,D.H., Brunschweiler,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, São Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/ICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?lt1-PM4&ct2-PM4-ET0154251100-00-d09&ct3-2000-11-25&ct4-1>)
Seq primer: puc 18 forward
High quality sequence start: 33
High quality sequence stop: 362.

```

FEATURES
SOURCE
LOCATION/Qualifiers
1. .362
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="E0154"
/dev_stage="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site.1: SmaI;
Site.2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT
129 a 74 c 108 g 51 t
ORIGIN

```

Query Match	11.3%;	Score 35.6;	DB 12;	Length 362;
Best Local Similarity	47.7%;	Pred. No. 4;		
Matches 104;	Conservative	0;	Mismatches 114;	Indels 0; Gaps 0
QY	51	CCCTTATAGGCCCTACACCCCTGCTCTTTGGCCCTGTAGTACTGTGGCTGAGTAT	110	
Db	270	CCGTGGTGGAAATCCTCATCTCATCTTCTTCTCCCTGATTAATGATTTGGTTTT	211	
QY	111	GAGCGGCACCTCCCAAGACATGGCTGGAACAATGTCTCTGCTCCCTCCACACAGCATGAC	170	
Db	210	GGACTCTCCCTTACTATACATCTCCCTTCTCTCTCTCTCTCTTCCATCATGCTCTCTTGGGT	151	
QY	171	AGATTTCACCTGGGGAGCTGGGGGCAAGAAGACCTGTGCACAAACAAAAGAGCCAGG	230	
Db	150	GGAGATTACGGGTAGCGTACGAGCTTGTGGGAGCTGTGCTGAACATCTTGTGGGCTATC	91	
QY	231	TCACACATGCTGATACCTCTGCCCTGCCCTCACCCATGGC	268	
Db	90	TCCTCAGGTGTGGGCATGTATCTCATCTCTTAATGTC	53	

RESULT 6					
BM030759/c	BM030759	389 bp	mRNA	linear	EST 05-NOV-2001
LOCUS					
DEFINITION	495261 MARC 2BOV Bos	taurus	CDNA 5',	mrna	sequence.
ACCESSION	BM030759				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Euarystota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1088)	Genoscope.	Direct Submission		
Submitted (23-JUL-1999)	Genoscope - Centre National de Sequencage			
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr				
- Web : www.genoscope.cns.fr)				
Determination of this BAC-end sequence was carried out as part of a				
collaboration with the European Drosophila Genome Project (EDGP) -				
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC				
library (Dros BAC) was made by Alain Billard at CEPH (Centre				
d'Etude du Polymorphisme Humain) with funding provided by a MRC				
project grant. The DNA was prepared from embryos by Alain Bucheton				
and Genevieve Payan. It has been constructed in the vector				
pBelorBAC11.				
FEATURES				
source				
1..1088				
/organism="Drosophila melanogaster"				
/db_xref="taxon:7227"				
/clone="BACN07J08"				
/clone_lib="DrosBAC"				
/plasmid="pBelorBAC11"				
/note="end : 77"				
BASE COUNT				
233 a 185 c 103 g 108 t 459 others				
Query Match				
Best Local Similarity				
Matches				
37; Conservative				
84; Mismatches				
83; Indels				
0; Gaps				
0;				
QY				
65				
ACCGTCGCCCTTCTTGGCCCTTAGTACTGTGGCGGAGATGACGGCCACTGCCA				
124				
DB				
1087				
MCNSKCKMCCDDCKBKKCKBCKBCKMCKMCKMCKMCKMCKMCKMCKMCKMCKMCK				
1028				
QY				
125				
AGCATGGCGCTTGGAACAATGTCCTCTCGCCCTCCACAGACAGACAGATTCACTGGG				
184				
DB				
1027				
BNNCKKMKCKKCKBCKMCCCKKCKBCKMCCCKKCKBCKMCCCKKCKBCKMCCCKK				
968				
QY				
185				
ACTCGGGGGGCAAGAGAGACTGTGTGCACAGACAGAAAGAGCCAGGGTCACATGTGTA				
244				
DB				
967				
MGGGCVVAGGGGGGAGAGAMCGMAGMAGMAGMAGMAGMAGMAGMAGMAGMAGMAG				
908				
QY				
245				
CCCTCGGCCCTGCTCACCCATGCC				
268				
DB				
907				
KMCKBKMKACGCMCCMCKCKKCK				
884				
RESULT 9				
BI823340				
LOCUS				
DEFINITION				
603041251F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5182069 5',				
mRNA sequence.				
BI823340				
VERSION				
BI823340.1				
GI:15934890				
EST.				
KEYWORDS				
human.				
ORGANISM				
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
1 (bases 1 to 1079)				
NIH-MGC http://mgc.nci.nih.gov/				
National Institutes of Health, Mammalian Gene Collection (MGC)				
Unpublished (1999)				
Contact: Robert Strausberg, Ph.D.				
Email: cga@bbs.fda.gov				
Tissue Procurement: Life Technologies, Inc.				
cDNA Library Preparation: Life Technologies, Inc.				
DNA Sequencing by: Incyte Genomics, Inc.				
Clone distribution: MGC clone distribution information can be				
found through the I.M.A.G.E. Consortium/ILNU at:				
http://image.llnl.gov				
Plate: U1A11454 row: e column: 14				
High quality sequence stop: 355.				

```
FEATURES                                     Location/Qualifiers
SOURCE                                         1..1079
                                                /organism="Homo sapiens"
                                                /db_xref="taxon:9606"
                                                /clone="IMAGE:5182069"
                                                /clone_1lb="NIH_MGC_115"
                                                /lab_host="DHIOB"
        note= "Organ: pooled brain, lung, testis; Vector: pCMV-Sport6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, Insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen) . Research Genetics tracking code 021. Note: this is a NIH-MGC library."

BASE COUNT      207 a   307 c   412 g   153 t

ORIGIN
Query Match          11.0%; Score 34.6; DB 13; Length 1079;
Best Local Similarity 50.3%; Pred. No. 12;
Matches            85; Conservative    0; Mismatches     84; Indels       0; Gaps         0;

OY      66 CCCGCCCCCTTTCTTGCCGCCCTGTGAAGTAGTCGTGGCTGGACTATGACGGCACCATTCCCACA 125
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      251 CCAGATCGCCGCTCTCGCTCACTGTCACGCTCTGTGTGGCAGCATGATCGAAGCGCGCACAA 310
OY      126 GCATGGCCCTGTAACAATAAGTCTCCTCGGCCCTCCACACAGCAGTAGCATTCACCTGGGAGA 185
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      311 GGCTGCGCTGCAABATATCCCGACTGTGGCGGCGGTGAGCGCGCTGCCAAGGCCCCACGGCGG 370
OY      186 CTCGGGGGCAAAGAGAGACTGTGTCCAACAAGAACMAAGAGCCAGGGGTAC 234
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      371 CTGGGGGCCAACGGGCCCAAGAGAGGTGCACGCTCTGAAGAGACTGAGGGAC 419

RESULT 10
AI800420/c
LOCUS
DEFINITION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homosapiens
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT
Contact: Robert Strausberg, Ph.D.
Email: cgapssrf@mail.nih.gov
Tissue Procurement: Christopher Moskalkov, M.D., Ph.D., Michael R. Emmer-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone information can be found through the I.M.A.G.E Consortium/LMW at: www.bio.lnlnl.gov/bdrr/image/image.html
Insert length: 1680 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 416.
Location/qualifiers
1..518
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2141467"
/clone_1lb="NCI_CGAP_Gas4"
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2003, 03:55:38 ; Search time 112 Seconds
(without alignments)
4127.139 Million cell updates/sec

Title: US-09-955-807-1_COPY_59_373

Perfect score: 315

Sequence: 1 atcgtggtatctcagcc.....atgtgtcttcagcgagccac 315

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NIM_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NIM_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	315	100.0	1692	US-09-955-807-1	Sequence 1, Appl
2	313.4	99.5	12001	US-09-955-807-13	Sequence 13, Appl
3	304	96.5	384	US-09-955-807-5	Sequence 6, Appl
4	36.2	11.5	3736	US-09-955-830-10	Sequence 10, Appl
5	33.8	10.7	414	US-10-123-155-418	Sequence 418, Appl
6	33.4	10.6	98865	US-09-770-689A-3	Sequence 3, Appl
7	32.6	10.3	81826	US-10-175-523-197	Sequence 197, Appl
8	32.4	10.3	822	US-09-893-737-293	Sequence 293, Appl
9	32.4	10.3	3428	US-09-822-846-304	Sequence 204, Appl
10	31.4	10.0	3012	US-10-151-542A-18	Sequence 18, Appl
11	31	9.8	2762	US-10-028-072-13	Sequence 13, Appl
12	31	9.8	2762	US-10-121-049-13	Sequence 13, Appl
13	31	9.8	2762	US-10-123-904-13	Sequence 13, Appl
14	31	9.8	2762	US-10-140-470-13	Sequence 13, Appl
15	31	9.8	2762	US-10-175-746-13	Sequence 13, Appl
16	31	9.8	2762	US-10-176-918-13	Sequence 13, Appl
17	31	9.8	2762	US-10-176-921-13	Sequence 13, Appl
18	31	9.8	2762	US-10-137-865-13	Sequence 13, Appl
19	31	9.8	2762	US-10-140-474-13	Sequence 13, Appl

C 20	31	9.8	2762	9	US-10-142-431-13	Sequence 13, Appl
C 21	31	9.8	2762	9	US-10-143-114-13	Sequence 13, Appl
C 22	31	9.8	2762	9	US-10-140-002-13	Sequence 13, Appl
C 23	31	9.8	2762	9	US-10-142-419-13	Sequence 13, Appl
C 24	31	9.8	2762	9	US-10-123-262-13	Sequence 13, Appl
C 25	31	9.8	2762	9	US-10-142-423-13	Sequence 13, Appl
C 26	31	9.8	2762	9	US-10-121-050-13	Sequence 13, Appl
C 27	31	9.8	2762	9	US-10-141-755-13	Sequence 13, Appl
C 28	31	9.8	2762	9	US-10-143-032-13	Sequence 13, Appl
C 29	31	9.8	2762	9	US-10-123-108-13	Sequence 13, Appl
C 30	31	9.8	2762	9	US-10-123-236-13	Sequence 13, Appl
C 31	31	9.8	2762	9	US-10-123-261-13	Sequence 13, Appl
C 32	31	9.8	2762	9	US-10-140-921-13	Sequence 13, Appl
C 33	31	9.8	2762	9	US-10-140-928-13	Sequence 13, Appl
C 34	31	9.8	2762	9	US-10-121-045-13	Sequence 13, Appl
C 35	31	9.8	2762	9	US-10-123-292-13	Sequence 13, Appl
C 36	31	9.8	2762	9	US-10-123-903-13	Sequence 13, Appl
C 37	31	9.8	2762	9	US-10-124-819-13	Sequence 13, Appl
C 38	31	9.8	2762	9	US-10-124-822-13	Sequence 13, Appl
C 39	31	9.8	2762	9	US-10-140-925-13	Sequence 13, Appl
C 40	31	9.8	2762	9	US-10-160-498-13	Sequence 13, Appl
C 41	31	9.8	2762	9	US-10-121-041-13	Sequence 13, Appl
C 42	31	9.8	2762	9	US-10-121-043-13	Sequence 13, Appl
C 43	31	9.8	2762	9	US-10-121-047-13	Sequence 13, Appl
C 44	31	9.8	2762	9	US-10-123-215-13	Sequence 13, Appl
C 45	31	9.8	2762	9	US-10-123-902-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-955-807-1
Sequence 1, Application US/09955807
Patent No. US20020132996A1
GENERAL INFORMATION:
APPLICANT: Lok, Si
APPLICANT: Sheppard, Paul O.
APPLICANT: Kindsvogel, Wayne
APPLICANT: Bort, Susan J.
TITLE OF INVENTION: Secretory Protein-48
FILE REFERENCE: 98-17C1
CURRENT APPLICATION NUMBER: US/09/955,807
CURRENT FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: 60/102,679
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 09/410,603
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1692
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (59)...(373)
US-09-955-807-1
Query Match 100.0%; Score 315; DB 10; Length 1692;
Best Local Similarity 100.0%; Pred. No. 4.2e-96;
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ANGCGGGTATTTCGAGCCCATGCGCATGTGCACACCCACTGGCTCTCTCTTACGC 60
|||||
DB 59 ANGCGGGTATTTCGAGCCCATGCGCATGTGCACACCCACTGGCTCTCTCTTACGC 118
OY 61 CACACCCGCGCCCTTCTTGGCCCTGTAGTACTGTGGTGGATGAGTACGAGCCACT 120
|||||
DB 119 CTACACCCGCGCCCTTCTTGGCCCTGTAGTACTGTGGTGGATGAGTACGAGCCACT 178
OY 121 CCCAGCATGGCGCTGGACATATGCTCCTGCGCCCTCCACACGAGTACAGATTCAC 180
|||||

Db 179 CCCAAGCANTGGCTGGAACAATGTCTCTCCGCCCTCCACAGCAGTGTACAGATTGACT 238
QY 181 GGGGACTCGGGGCAAGAGAGACTGTGTCAACAAGACAAAGAGACCGAGGTACACATGG 240
Db 239 GGGGACTCGGGGCAAGAGAGACTGTGTCAACAAGACAAAGAGACCGAGGTACACATGG 298
QY 241 TGTACCTCGCCCTGCTCTACCCATGGCTGACATGGTTGGACACCTCAGAAATCATGTG 300
Db 299 TGTACCTCGCCCTGCTCTACCCATGGCTGACATGGTTGGACACCTCAGAAATCATGTG 358
QY 301 TCTTCAGCGAGCCAC 315
Db 359 TCTTCAGCGAGCCAC 373

RESULT 2

US-09-955-807-13
; Sequence 13, Application US/09955807
; Patent No. US20020132996A1
; GENERAL INFORMATION:
; APPLICANT: Lok, SI
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Bort, Susan J.
; TITLE OF INVENTION: Secretory Protein-48
; FILE REFERENCE: 98-17C1
; CURRENT APPLICATION NUMBER: US/09/955, 807
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/102,679
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 09/410,603
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 12001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10258)...(10572)
US-09-955-807-13

Query Match 99.5%; Score 313.4; DB 10; Length 12001;
Best Local Similarity 99.7%; Pred. No. 2,5e-95;
Matches 314; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCTGGTATTTCTGAGCCATGCGCATGTGACACACCCACTTGGCCCTCTCCTTTAGGC 60
Db 10258 ATGCTGGTATTTCTGAGCCATGCGCATGTGACACACCCACTTGGCCCTCTCCTTTAGGC 10317
QY 61 CTACACCCCTGCTCTTTCTTTGCCCCCTTGTAGTTACTGTGCTGAGTGTAGTACGCCACT 120
Db 10318 CTACACCCCTGCTCTTTCTTTGCCCCCTTGTAGTTACTGTGCTGAGTGTAGTACGCCACT 10377
QY 121 CCCAAGCATGGCTGTGAACAATGTCTCTGCCCCCTCCACAGCATGTGACAGATTGACT 180
Db 10378 CCCAAGCATGGCTGTGAACAATGTCTCTGCCCCCTCCACAGCATGTGACAGATTGACT 10437
QY 181 GGGGACTCGGGGCAAGAGAGACTGTGTCAACAAGACAAAGAGACCGAGGTACACATGG 240
Db 10438 GGGGACTCGGGGCAAGAGAGACTGTGTCAACAAGACAAAGAGACCGAGGTACACATGG 10497
QY 241 TGTACCTCGCCCTGCTCTACCCATGGCTGACATGGTTGGACACCTCAGAAATCATGTG 300
Db 10498 TGTACCTCGCCCTGCTCTACCCATGGCTGACATGGTTGGACACCTCAGAAATCATGTG 10557
QY 301 TCTTCAGCGAGCCAC 315
Db 10558 TCTTCAGCGAGCCAC 10572

RESULT 3

US-09-955-807-6
; Sequence 6, Application US/09955807
; Patent No. US20020132996A1
; GENERAL INFORMATION:
; APPLICANT: Lok, SI
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Bort, Susan J.
; TITLE OF INVENTION: Secretory Protein-48
; FILE REFERENCE: 98-17C1
; CURRENT APPLICATION NUMBER: US/09/955, 807
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/102,679
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 09/410,603
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 384
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(384)
; OTHER INFORMATION: n is any nucleotide
; NAME/KEY: misc-feature
; LOCATION: (1)...(384)
; OTHER INFORMATION: n = A,T,C or G
US-09-955-807-6

Query Match 96.5%; Score 304; DB 10; Length 384;
Best Local Similarity 97.5%; Pred. No. 1.4e-92;
Matches 307; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGCTGGTATTTCTGAGCCATGCGCATGTGACACACCCACTTGGCCCTCTCCTTTAGGC 60
Db 59 ATGCTGGTATTTCTGAGCCATGCGCATGTGACACACCCACTTGGCCCTCTCCTTTAGGC 118
QY 61 CTACACCCCTGCTCTTTCTTTGCCCCCTTGTAGTTACTGTGCTGAGTGTAGTACGCCACT 120
Db 119 CTACACCCCTGCTCTTTCTTTGCCCCCTTGTAGTTACTGTGCTGAGTGTAGTACGCCACT 178
QY 121 CCCAAGCATGGCTGTGAACAATGTCTCTGCCCCCTCCACAGCATGTGACAGATTGACT 180
Db 179 CCCAAGCATGGCTGTGAACAATGTCTCTGCCCCCTCCACAGCATGTGACAGATTGACT 238
QY 181 GGGGACTCGGGGCAAGAGAGACTGTGTCAACAAGACAAAGAGACCGAGGTACACATGG 240
Db 239 GGGGACTCGGGGCAAGAGAGACTGTGTCAACAAGACAAAGAGACCGAGGTACACATGG 298
QY 241 TGTACCTCGCCCTGCTCTACCCATGGCTGACATGGTTGGACACCTCAGAAATCATGTG 300
Db 299 TGTACCTCGCCCTGCTCTACCCATGGCTGACATGGTTGGACACCTCAGAAATCATGTG 358
QY 301 TCTTCAGCGAGCCAC 315
Db 359 TCTTCAGCGAGCCAC 373

RESULT 4

US-09-965-830-10
; Sequence 10, Application US/09965830
; Patent No. US20020177201A1
; GENERAL INFORMATION:
; APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.
; TITLE OF INVENTION: A novel potassium channel protein
; FILE REFERENCE: 19903-PCT
; CURRENT APPLICATION NUMBER: US/09/965, 830
; CURRENT FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 09/600,776
; PRIOR FILING DATE: 2001-07-21
; PRIOR APPLICATION NUMBER: JP P1998-346198


```

; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 09/581,033
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: PCT/US98/26422
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 60/069,379
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 18
; LENGTH: 3012
; TYPE: DNA
; ORGANISM: Homo sapien (human)
US-10-151-542A-18

Query Match      10.0%; Score 31.4; DB 9; Length 3012;
Best Local Similarity 49.1%; Pred. No. 1.7;
Matches 83; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY      63 ACACCTGCCCTTCTTGGCCCTGTAGTCTGTGCTGAGTGAATGAGGCCACTCC 122
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      1247 AGACCTGGGGCTTGAAGTCCCTAGCCATCATCTCTAGTGTGCACCCGACCAT 1188
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      123 CAAGCATGGCTGGAACATGTCTCTGCCCCCTCCACGAGTGAACAGATTACTGG 182
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      1187 CATCCCTGGCCAGATGATGACACCCGACCCGACCATCTCGAGATGACCCGACG 1128
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      183 GGATCGGGGGGCAAGGAGACTGTGTCAACAGACAAAGAGCCAGGGT 231
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      1127 CATCCCTGAATCAGAGTATGTGTCTCCGACCATCATCAACCAAGCT 1079
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 11
US-10-028-072-13/C
; Sequence 13, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
```

```

; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059836
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062285
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062814
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/062816
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063082
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/063127
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063327
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063329
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063550
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063561
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063704
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063733
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063735
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063738
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064248
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/064809
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065846
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066453
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069212
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069278
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069334
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069694
; PRIOR FILING DATE: 1997-12-16
; PRIOR APPLICATION NUMBER: 60/072320
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: 60/073612
; PRIOR FILING DATE: 1998-02-04
; PRIOR APPLICATION NUMBER: 60/074086
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/074092
; PRIOR FILING DATE: 1998-02-09
```

;; PRIOR APPLICATION NUMBER: 60/077791
;; PRIOR FILING DATE: 1998-03-12
;; PRIOR APPLICATION NUMBER: 60/078910
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/079294
;; PRIOR FILING DATE: 1998-03-25
;; PRIOR APPLICATION NUMBER: 60/079663
;; PRIOR FILING DATE: 1998-02-27
;; PRIOR APPLICATION NUMBER: 60/079728
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/080165
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/081203
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081229
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081695
;; PRIOR FILING DATE: 1998-04-14
;; PRIOR APPLICATION NUMBER: 60/081817
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081818
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/082999
;; PRIOR FILING DATE: 1998-04-24
;; PRIOR APPLICATION NUMBER: 60/083322
;; PRIOR FILING DATE: 1998-04-28
;; PRIOR APPLICATION NUMBER: 60/083545
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/084600
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084627
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084637
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085149
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085323
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085338
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085339
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085697
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085704
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/086414
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/086430
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087106
;; PRIOR FILING DATE: 1998-05-28
;; PRIOR APPLICATION NUMBER: 60/088026
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088730
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088741
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088810
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088858
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089532
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089599
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089907
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089947
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/090349

;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090429
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090538
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07

Query Match 9.8%; Score 31; DB 9; Length 2762;
Best Local Similarity 62.0%; Pred. No. 2.2;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 80 TGGCCCTTGTAGTACTGTGCTGAGCTGATGAGCGCCACTCCCAAGCATGCGCTGGAC 139
Db 1973 TCCCACTCTAGTATGAGGGGCTGGGTATGAGAGACCGTCCACGCGGGCGCTGACG 1914

QY 140 AATGTCTCTCGCCCTCC 158
Db 1913 CTCCTACTCTCTCGCCCTCC 1895

RESULT 12
US-10-121-049-13/c

;; Sequence 13, Application US/10121049
;; Publication No. US2003002239A1
;; GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Beresini, Maureen
;; APPLICANT: Deforge, Laura
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Filvaert, Ellen
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerlitsen, Mary E.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Sherwood, Steven
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Watanabe, Colin K
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; TITLE OF INVENTION: ACIDS ENCODING THE SAME
;; FILE REFERENCE: P3330R1C17
;; CURRENT APPLICATION NUMBER: US/10/121,049
;; CURRENT FILING DATE: 2002-04-12
;; PRIOR APPLICATION removed - See File Wrapper or Palm
;; NUMBER OF SEQ ID NOS: 550
;; SEQ ID NO 13
;; LENGTH: 2762
;; TYPE: DNA
;; ORGANISM: Homo Sapien
US-10-121-049-13

Query Match 9.8%; Score 31; DB 9; Length 2762;
Best Local Similarity 62.0%; Pred. No. 2.2;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 80 TGGCCCTTGTAGTACTGTGCTGAGCTGATGAGCGCCACTCCCAAGCATGCGCTGGAC 139
Db 1973 TCCCACTCTAGTATGAGGGGCTGGGTATGAGAGACCGTCCACGCGGGCGCTGACG 1914
QY 140 AATGTCTCTCTCGCCCTCC 158

DB 1913 CTCCTACTCTGCGGCTCC 1895

RESULT 13

US-10-123-904-13/c
Sequence 13, Application US/10123904
Publication No. US2003002238A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P330R1C54
CURRENT APPLICATION NUMBER: US/10/123,904
CURRENT FILING DATE: 2002-04-16
Prior Application removed - See File Wrapper or Palm
SEQ ID NO 13
LENGTH: 2762
TYPE: DNA
ORGANISM: Homo Sapien
US-10-123-904-13

Query Match

Best Local Similarity 9.8%; Score 31; DB 9; Length 2762;
Best Local Similarity 62.0%; Pred. No. 2.2;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 80 TGCCCCCTGTAGTACTGTGCTGGAGTATGAGAGCCGACCTCCAGCATGGCCCTGGAGAC 139

DB 1973 TCCCCACTCTAGTTAGAGGGGCTGGGATGAGAGACCGTCCAGCCGGGCGCTGCACG 1914

QY 140 AATGTCTCTCTGCCCCCTCC 158

DB 1913 CTCCTACTCTGCGGCTCC 1895

RESULT 14

US-10-140-470-13/c
Sequence 13, Application US/10140470
Publication No. US2003002233A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William

APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P330R1C160
CURRENT APPLICATION NUMBER: US/10/140,470
CURRENT FILING DATE: 2002-05-06
Prior Application removed - See File Wrapper
SEQ ID NO 13
LENGTH: 2762
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-470-13

Query Match 9.8%; Score 31; DB 9; Length 2762;
Best Local Similarity 62.0%; Pred. No. 2.2;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 80 TGCCCCCTGTAGTACTGTGCTGGAGTATGAGAGCCGACCTCCCAAGCATGGCCCTGGAGAC 139

DB 1973 TCCCCACTCTAGTTAGAGGGGCTGGGATGAGAGACCGTCCAGCCGGGCGCTGCACG 1914

QY 140 AATGTCTCTCTGCCCCCTCC 158

DB 1913 CTCCTACTCTGCGGCTCC 1895

QY 80 TGCCCCCTGTAGTACTGTGCTGGAGTATGAGAGCCGACCTCCCAAGCATGGCCCTGGAGAC 139

DB 1973 TCCCCACTCTAGTTAGAGGGGCTGGGATGAGAGACCGTCCAGCCGGGCGCTGCACG 1914

QY 140 AATGTCTCTCTGCCCCCTCC 158

DB 1913 CTCCTACTCTGCGGCTCC 1895

RESULT 15

US-10-175-746-13/c
Sequence 13, Application US/10175746
Publication No. US20030027270A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P330R1C353
CURRENT APPLICATION NUMBER: US/10/175,746
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 13
LENGTH: 2762
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-746-13

Query Match

Best Local Similarity 9.8%; Score 31; DB 9; Length 2762;
Best Local Similarity 62.0%; Pred. No. 2.2;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 80 TGCCCCCTGTAGTACTGTGCTGGAGTATGAGAGCCGACCTCCCAAGCATGGCCCTGGAGAC 139

DB 1973 TCCCCACTCTAGTTAGAGGGGCTGGGATGAGAGACCGTCCAGCCGGGCGCTGCACG 1914

QY 140 AATGTCTCTCTGCCCCCTCC 158

DB 1913 CTCCTACTCTGCGGCTCC 1895

Mon Jun 23 10:42:32 2003

us-09-955-807-1_copy_59_373.rnpb

Page 8

Search completed: June 21, 2003, 04:55:35
Job time : 114 secs

APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 10.5%; Score 33; DB 4; Length 4403765;
Best Local Similarity 55.8%; Pred. No. 8.3;
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 33 ACACCCACTTGGGCTCTCTCTTAGGCTACACCCCTTCTTCTTCCCTTGTAGT 92
DB 1838235 ACCCCGACGAGGCTCTCTCTTCCGCGCAACCACTGAGATGTTGGTCTGTGC 1838294

QY 93 TACTGTGGCTGAGTGAAGCCCACTCCCAAGATGGCTGAGCAATGTC 145
DB 1838295 CACCGTTCCGGTGAATAATCACCGCGCATGTGACAAACGGCTAATACAGTGTG 1838347

RESULT 3
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 10.5%; Score 33; DB 4; Length 4411529;
Best Local Similarity 55.8%; Pred. No. 8.3;
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 33 ACACCCACTTGGGCTCTCTCTTAGGCTACACCCCTTCTTCTTCCCTTGTAGT 92
DB 1847358 ACCCCGACGAGGCTCTCTTCCGCGCAACCACTGAGATGTTGGTCTGTGC 1847417

QY 93 TACTGTGGCTGAGTGAAGCCCACTCCCAAGATGGCTGAGCAATGTC 145
DB 1847418 CACCGTTCCGGTGAATAATCACCGCGCATGTGACAAACGGCTAATACAGTGTG 1847470

RESULT 4
US-09-060-756-295
; Sequence 295, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:

APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 295
LENGTH: 321
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-060-756-295

Query Match 9.8%; Score 31; DB 4; Length 321;
Best Local Similarity 55.0%; Pred. No. 0.85;
Matches 61; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 4 CTGGTATCTAGCCCATGCGACACCACTTGGCTCTCTTAGGCTTA 63
DB 120 CGGTGATCATCAGAGGCTTGGTCAGTGGGACACCACTACGCAATTTCTCCGACTGGGA 179

QY 64 CACCTGCGCCCTTCTTGGCCCTTGTAGTACTGTGAGTGAATGAGC 114
DB 180 CACCTACCGAGGCTCGCCCACTGCAAGAGTGTGTCCCGCAACGGGC 230

RESULT 5
US-08-146-930-3
; Sequence 3, Application US/08146930
; Patent No. 5958764
; GENERAL INFORMATION:
; APPLICANT: ROOP, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; TITLE OF INVENTION: SPECIFIC EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: Mordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/146,930
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; APPLICATION DATA: described below: two
; FILING DATE: April 30, 1992
; APPLICATION NUMBER: 07/876,286
; APPLICATION NUMBER: No. 5958764 yet assigned (204/132)
; FILING DATE: October 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 204/152
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510


```
APPLICANT: Basilio, Claudio
TITLE OF INVENTION: NOVEL GROWTH FACTOR INDUCIBLE
SERINE/THREONINE PHOSPHATASE, FIN13
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
FLOOR
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,701
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1049-1-002 N
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1824 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-822-701-1

Query Match
Best Local Similarity 9.4%; Score 29.6; DB 2; Length 1824;
Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 58 GGCCTACACCCCTCTTCTTGGCCCTGTAGTACTGTGCTGGAGTATGATGACGCC 117
DB 1401 GGTTCACACCTAGTCCCTCTTGGCCCTGTAGTACTGTGCTGGAGTATGATGACGCC 1342
QY 118 ACTCCAGACAGTGGCTGGAACAATGTCTCTGCCCCC 157
DB 1341 CTCCTGACAGTGCCTCTCCAGTTCTCTTGCACATC 1302

RESULT 12
US-08-935-855-1/c
Sequence 1, Application US/08935855
Patent No. 6066485
GENERAL INFORMATION:
APPLICANT: Guthridge, Mark
APPLICANT: Basilio, Claudio
TITLE OF INVENTION: NOVEL GROWTH FACTOR INDUCIBLE
SERINE/THREONINE PHOSPHATASE, FIN13
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESS: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
FLOOR
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/935,855
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1049-1-002 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1824 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-935-855-1

Query Match
Best Local Similarity 9.4%; Score 29.6; DB 3; Length 1824;
Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 58 GGCCTACACCCCTCTTCTTGGCCCTGTAGTACTGTGCTGGAGTATGATGACGCC 117
DB 1401 GGTTCACACCTAGTCCCTCTTGGCCCTGTAGTACTGTGCTGGAGTATGATGACGCC 1342
QY 118 ACTCCAGACAGTGGCTGGAACAATGTCTCTGCCCCC 157
DB 1341 CTCCTGACAGTGCCTCTCCAGTTCTCTTGCACATC 1302

RESULT 13
US-08-935-855-21/c
Sequence 21, Application US/08935855
Patent No. 6066485
GENERAL INFORMATION:
APPLICANT: Guthridge, Mark
APPLICANT: Basilio, Claudio
TITLE OF INVENTION: NOVEL GROWTH FACTOR INDUCIBLE
SERINE/THREONINE PHOSPHATASE, FIN13
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESS: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
FLOOR
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/935,855
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1049-1-002 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
```

```

? LENGTH: 2081 base pairs
?
? TYPE: nucleic acid
?
? STRANDEDNESS: double
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: cDNA
?
? HYPOTHETICAL: NO
?
? ORIGINAL SOURCE:
?
? ORGANISM: Mus musculus
?
US-08-935-855-21

```

Query Match	9.48;	Score 29.6;	DB 3;	Length 2081;
Best Local Similarity	56.0%;	Pred. No. 5.1;		
Matches	56;	Conservative	0;	Mismatches 44;
				Indels 0

```

? Patient No. 6033871
?
? CURRENT APPLICATION NUMBER: US/08/922.635A
?
? CURRENT FILING DATE: 1997-09-03
?
? EARLIER APPLICATION NUMBER: 08/650.766
?
? EARLIER FILING DATE: 1996-05-20
?
? EARLIER APPLICATION NUMBER: 60/012,600
?
? EARLIER FILING DATE: 1996-03-01
?
? NUMBER OF SEQ ID NOS: 22
?
? SOFTWARE: PatentIn Ver. 2.0
?
? SEQ ID NO 21
?
? LENGTH: 15202
?
? TYPE: DNA
?
? ORGANISM: Homo sapiens

```

QY 58 GGGCTACACCCCTGGCCCTTTTCTTTCGCCCTTGATTTACTGTGGGTGATGATGAGGCC 117

Db 1658 GGTGACACCACTAGCCCTCTTGCCCTCTTTTTCACCTGTTGGCCGGTGTTCAGCCCC 1599

QY 118 ACTGCCAAGCATGGCCCTGGAAACAATGTCGTCCTGCCCTC 157

Db 1598 CTCGTGTGACACGTGCTCCCTCCACGATTCCTTTGCCACTC 1559

Query Match	9.4%	Score	29.6	DB	3	Length	15202
Best Local Similarity	54.6%	Pred. No.	11				
Matches	59	Conservative	0	Mismatches	49	Indels	0
						Gaps	0

RESULT 14
US-08-894-997-49/c
; Sequence 49, Application US/08894997A
; Patent No. 6270990

Search completed: June 21, 2003, 03:56:25
Job time : 46 secs

```

? APPLICANT: Schoenherz, Christopher J
? TITLE OF INVENTION: NEURON-RESTRICTIVE SILENCER FACTOR
? FILE REFERENCE: 17810-502 NRSF
? CURRENT APPLICATION NUMBER: US/08/894,997A
? EARLIER FILING DATE: 1998-01-06
? EARLIER APPLICATION NUMBER: PCT/US96/02817
? EARLIER FILING DATE: 1996-03-01
? EARLIER APPLICATION NUMBER: 08/398,550
? EARLIER FILING DATE: 1995-03-03
? NUMBER OF SEQ ID NOS: 55
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 49
? LENGTH: 4057
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: gene
? LOCATION: (1)..(4057)
? OTHER INFORMATION: Human NSRF
? JS-08-894-997-49

```

Oy	16	GAGCCCATGCGATGTGCACACCACCACTTGGCCTTCCTTGACTTAAGGCGTTACACCTTGCCTT	75
Dd	8372	GGGCCCCCTCCAGCATGGCACTGACACTGCGCTCTCCCTGCACAGGCCAAGAACATATGCCAG	8433
Oy	76	TCTTTGCCCTTGTAGTATTACTGTGTGGCTGGAGATGATGAAGCCCACTCCC	123
	8432	CCGAGGATCCAGGTGTGTGCCGGGGTGTGGCCAGATCATCTTCTGCCCC	8479

Query Match	9.44:	Score 29.6:	DB 4:	Length 4057:
Best local	Similarity	64.7%:	Pred. No. 6.6:	
Matches	44:	Conservative	0:	Mismatches 24: Indels 0: Gaps 0:

Oy

212 AAGACAAAGAAGCCAGGGGTGACATCGTCTGCCCGCCTAACCCCATGGGCATA **271**

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

D8 AAGCCCAGAGGCACTCTGAACAATTCTCTGTGATCTCCAACCCTTTGGCATGGGCITA **2759**

Qy	272	CATGGGT	279
		—	
Db	2758	GTTTGT	2751

RESULT 15
US-08-922

; Sequence 21, Application US/08922635A
; Patent No. 6033871

```

; GENERAL INFORMATION:
; APPLICANT: PILETZ, John E.
; APPLICANT: IVANOV, Tilda R.
; TITLE OF INVENTION: DNA MOLECULES ENCODING IMIDALINE RECEPTIVE POLYPEPTIDES
; TITLE OF INVENTION: AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: Corrected Sequence Listing

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2003, 03:25:03 ; Search time 157 Seconds
(without alignments)
4518.341 Million cell updates/sec

Title: US-09-955-807-1_COPY_59_373

Perfect score: 315
Sequence: 1 atgtcttggtattcttgagcc.....atgtcttcttcagcgagccac 315

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_101002:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	315	100.0	1692	21	AAZ94889 Human secretory pr
2	313.4	99.5	12001	21	AAZ94892 Human secretory pr
3	304	96.5	384	21	AAZ94890 Human secretory pr
4	39.6	12.6	521	22	AAH97963 Marine 7-transmemb
5	36.2	11.5	3736	20	AAH84918 Rat brain specific
6	36.2	11.5	10732	21	AAH10594 Gene encoding a su
7	33	10.5	9545	22	AAK80111 Human immune/haema
8	33	10.5	4403765	22	AAI99683 Mycobacterium tube
9	33	10.5	4411529	22	AAI99682 Mycobacterium tube

10	32.6	10.3	378	24	ABK51611 DNA encoding human
11	32.6	10.3	2633	22	ABA20784 Human nervous syst
12	32.6	10.3	2643	14	AAQ39212 CENP-B CDNA. Homo
13	32.6	10.3	3717	21	AAA64660 DNA encoding centir
14	32.6	10.3	4712	24	ABK51618 Human DNA sequence
15	32.6	10.3	4748	24	AAAS94823 Human DNA sequence
16	32.4	10.3	322	24	ABK33136 DNA encoding novel
17	32.4	10.3	1125	20	AAK23941 Human calpain CAPN
18	32.4	10.3	3428	24	ABK35813 CDNA sequence #204
19	32.2	10.2	412	24	ABN22186 Human ORFX polynuc
20	32	10.2	6114	20	AAH83490 Human T-type volta
21	32	10.2	6132	20	AAH83489 Human T-type volta
22	31.8	10.1	3301	24	ABL89919 Human polynucleoti
23	31.8	10.1	5796	23	ABL20263 Drosophila melanog
24	31.8	10.1	7796	23	ABL20262 Drosophila melanog
25	31.8	10.1	23914	23	ABL20258 Drosophila melanog
26	31.6	10.0	361	23	ABV35643 Human prostate exp
27	31.6	10.0	361	23	ABV44449 Human prostate exp
28	31.6	10.0	700	22	AAV43303 DNA encoding novel
29	31.6	10.0	822	23	ABV14564 Human prostate exp
30	31.6	10.0	1840	22	ABA20783 Human nervous syst
31	31.4	10.0	568	22	AAH97966 Murine 7-transmemb
32	31.4	10.0	2089	23	ABL29841 Drosophila melanog
33	31.4	10.0	2177	22	AAF28358 Human BBSR PPAR cd
34	31.4	10.0	3012	20	AAV59096 Human nuclear rece
35	31.4	10.0	5151	23	ABL29840 Drosophila melanog
36	31.4	10.0	13395	22	AAK74325 Human immune/haema
37	31.2	9.9	484	22	AAI66350 Human polynucleoti
38	31.2	9.9	1188	21	AAH94816 Drosophila melanog
39	31.2	9.9	1188	21	AAV72255 Drosophila odorant
40	31.2	9.9	1191	23	ABL21499 Drosophila melanog
41	31.2	9.9	3471	23	ABL21498 Drosophila melanog
42	31	9.8	321	21	ABO62795 Mycobacterium tube
43	31	9.8	2409	24	ABQ31915 Human NF-kB activa
44	31	9.8	2630	24	ABN59640 Novel human coding
45	31	9.8	2681	22	AAK94372 Human full-length

ALIGNMENTS

RESULT 1	AAZ94889	AAZ94889 standard; CDNA, 1692 BP.
ID	AAZ94889	01-AUG-2000 (first entry)
AC	AAZ94889	Human secretory protein 48 (Zsig48) CDNA.
XX	XX	Secretory protein 48; Zsig48; human; leukocyte; proliferation; immunostimulant; adjuvant; immunomodulator; gene therapy;
XX	XX	chromosome 7q36.3; ss.
DE	XX	Homo sapiens.
XX	XX	Key
XX	XX	Location/Qualifiers
XX	XX	CD5
XX	XX	sig_peptide
XX	XX	mat_peptide
XX	XX	/*tag- a
XX	XX	/*tag- b
XX	XX	/*tag- c
XX	XX	/*note- "alternatively, the mature peptide is encoded by nucleotides 143..373, 179..373 or 134..373"
XX	XX	WO2000018796-A2.
XX	XX	06-APR-2000.
XX	XX	01-OCT-1999; 99WO-US22970.

XX 01-OCT-1998; 98US-0164740.
XX (ZYMO) ZYMOGENETICS INC.
XX
XX
XX Lok S. Sheppard PO.
XX WPI; 2000-303441/26.
XX P-PSDB; AAY79436.
XX
XX Human secretory protein 48 polypeptides and polynucleotides useful for
XX promoting leukocyte proliferation and for treating immunosuppressed
XX individuals -
XX
XX Example 1; Page 127-128; 141pp; English.
XX
XX The present sequence is that of cDNA coding for human secretory
XX protein 48 (Zs1948, see AAY79436), a protein that stimulates the
XX proliferation of peripheral blood mononuclear cells, i.e. T-cells,
XX B-cells and monocytes. An expressed sequence tag (see AA294892) was
XX isolated through the random sequencing of a mixed hematopoietic
XX cDNA library, and the present full-length cDNA was subsequently
XX obtained. The Zs1948 gene was mapped to human chromosome 7q36.3.
XX The invention provides Zs1948 polynucleotides, polypeptides,
XX antibodies and anti-idiotypic antibodies. Zs1948 stimulates the
XX proliferation of leukocytes in both a mixed leukocyte reaction and
XX in an unmixd leukocyte reaction. It can be used to promote
XX proliferation of peripheral blood leukocytes and to treat low
XX leukocyte counts in individuals (claimed). This can be useful in
XX treating cancer patients whose leukocytes have been depleted by
XX chemotherapy, radiation or illness. Zs1948 can be administered to
XX patients receiving bone marrow transplants to promote proliferation
XX of leukocytes produced by the transplanted marrow. It could also
XX be useful in treating immunosuppressed individuals such as the
XX elderly or HIV infected individuals, or used as a vaccine adjuvant.
XX Zs1948 polynucleotides can be used in the recombinant production of
XX Zs1948 polypeptides, as probes and primers to detect Zs1948 gene
XX expression or to identify mutations in the Zs1948 gene, in
XX diagnostic assays, or introduced into a subject to enhance
XX immunological responses (immunomodulator gene therapy).
XX
XX Sequence 1692 BP; 490 A; 393 C; 362 G; 447 T; 0 other;
SQ
Query Match 100.0%; Score 315; DB 21; Length 1692;
Best Local Similarity 100.0%; Pred. No. 2.1e-88;
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCTGGTATTTCTGAGCCCATGTCACACCCACTTGCCCTCTCTTAGGC 60
DB 59 ATGCTGGTATTTCTGAGCCCATGTCACACCCACTTGCCCTCTCTTAGGC 118
QY 61 CTACACCTGCTCTTCTTGGCCCTGTAGTACTGTGCTGAGATGAGCGCCACT 120
DB 119 CTACACCTGCTCTTCTTGGCCCTGTAGTACTGTGCTGAGATGAGCGCCACT 178
QY 121 CCCAAGCAGTGGCTGGAACATGTCTCTGCCCCCTCCACAGCAGTGCATGACT 180
DB 179 CCCAAGCAGTGGCTGGAACATGTCTCTGCCCCCTCCACAGCAGTGCATGACT 238
QY 181 GGGGACTCGGGGCAAGAGACTGTGTCAAGACAAAGAGAGCCAGGCTCACACTGG 240
DB 239 GGGGACTCGGGGCAAGAGACTGTGTCAAGACAAAGAGAGCCAGGCTCACACTGG 298
QY 241 TTAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
DB 299 TTAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 358
QY 301 TCTTACGAGGAGCCAC 315
DB 359 TCTTACGAGGAGCCAC 373
RESULT 2

AA294892
ID AA294892 standard; DNA; 12001 BP.
XX
XX AC AA294892;
XX
XX 01-AUG-2000 (first entry)
XX
XX Human secretory protein 48 (Zs1948) genomic DNA.
XX
XX
XX Secretory protein 48; Zs1948; human; leukocyte; proliferation;
XX immunostimulant; adjuvant; immunomodulator; gene therapy;
XX chromosome 7q36.3; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 10258..10575
XX FT /*tag= a
XX FT 10258..10335
XX FT /*tag= b
XX FT mat_peptide 10336..10572
XX FT /*tag= c
XX
XX W0200018796-A2.
XX
XX 06-APR-2000.
XX
XX 01-OCT-1999; 99WO-US22970.
XX
XX 01-OCT-1998; 98US-0164740.
XX (ZYMO) ZYMOGENETICS INC.
XX
XX Lok S. Sheppard PO.
XX WPI; 2000-303441/26.
XX P-PSDB; AAY79436.
XX
XX Human secretory protein 48 polypeptides and polynucleotides useful for
XX promoting leukocyte proliferation and for treating immunosuppressed
XX individuals -
XX
XX Disclosure; Page 132-137; 141pp; English.
XX
XX The present sequence is that of genomic DNA coding for human
XX secretory protein 48 (Zs1948, see AAY79436), a protein that stimulates
XX the proliferation of peripheral blood mononuclear cells, i.e. T-cells,
XX B-cells and monocytes. The gene maps to human chromosome 7q36.3.
XX The invention provides Zs1948 polynucleotides, polypeptides,
XX antibodies and anti-idiotypic antibodies. Zs1948 stimulates the
XX proliferation of leukocytes in both a mixed leukocyte reaction and
XX in an unmixd leukocyte reaction. It can be used to promote
XX proliferation of peripheral blood leukocytes and to treat low
XX leukocyte counts in individuals (claimed). This can be useful in
XX treating cancer patients whose leukocytes have been depleted by
XX chemotherapy, radiation or illness. Zs1948 can be administered to
XX patients receiving bone marrow transplants to promote proliferation
XX of leukocytes produced by the transplanted marrow. It could also
XX be useful in treating immunosuppressed individuals such as the
XX elderly or HIV infected individuals, or used as a vaccine adjuvant.
XX Zs1948 polynucleotides can be used in the recombinant production of
XX Zs1948 polypeptides, as probes and primers to detect Zs1948 gene
XX expression or to identify mutations in the Zs1948 gene, in
XX diagnostic assays, or introduced into a subject to enhance
XX immunological responses (immunomodulator gene therapy).
XX
XX Sequence 12001 BP; 3336 A; 2441 C; 2461 G; 3763 T; 0 other;
SQ
Query Match 99.5%; Score 313.4; DB 21; Length 12001;
Best Local Similarity 99.7%; Pred. No. 1.3e-87;
Matches 314; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCTGGTATTTCTGAGCCCATGTCACACCCACTTGCCCTCTCTTAGGC 60

PT DNA Sequences encoding 7-transmembrane G-protein coupled protein.
PT receptors characteristic of hematopoietic stem cells, useful for
PT treating leukemia -

XX
XX
PS Claim 1; Page 100; 176pp; English.

CC The present invention relates to murine coding sequences for
CC 7-transmembrane G-protein coupled protein receptors (7TM-GPCRs). The
CC present sequence is one such murine 7TM-GPCR coding sequence. The present
CC sequence was derived from hematopoietic stem cells. The present sequence
CC and its corresponding protein are useful in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate 7TM-GPCR expression.
CC 7TM-GPCRs identify specific signalling molecules, to activate an
CC effector-signalling cascade that triggers an intracellular response and
CC eventually a biological effect.

XX
SQ Sequence 521 BP; 100 A; 158 C; 143 G; 119 T; 1 other;

Query Match 12.6%; Score 39.6; DB 22; Length 521;
Best Local Similarity 46.4%; Pred. No. 0.021;
Matches 129; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 22 ATGCATGTGCACACCCACTTGGCTCTCTCTAGCCTACACCCCTGCTTCTTGG 81
DB 381 ATGTGTATCATCCTCCATACCCCTCTCCGCTCATCCCTCATCGGTGGCTTCTC 322
QY 82 CCCCTTGTACTTACTGTGCTGAGTATGAGCCCATCCCAAGCATGGCTGGAACA 141
DB 321 CTGTGAGGAGACAGCAGGATGAGAGACAGATGCGGCCAGATGAGAGCAGAGGA 262
QY 142 TGTCTCTCTCTCCCTCCACAGCAGTATGATTCATCTGGGACTCGGGGCAAGAG 201
DB 261 GGCACCTCCAGCCCAAGTAGAGGGGGGCCAGTTCATACCTGAGAGGAGAAAGTG 202
QY 202 ACTGTGTACAAAGACAAAGAGCAGGTATGATGATGATGATGATGATGATGATG 261
DB 201 AGTGTCCAGCCCAAGAGAGGTGTGGCTGAGACCCACCTGCTGTCTCCCAAGGGTTC 142
QY 262 CCATGGCTGACATGGGTGAGACACCTCAGAAATCATGT 299
DB 141 CACTTACTTGTGTTGCTGGAAGATCAGTAGATGT 104

RESULT 5
AAK84918
ID AAK84918 standard; DNA: 3736 BP.

XX
AC AAK84918;

XX
DT 28-SEP-1999 (first entry)

XX
DE Rat brain specific potassium channel protein coding sequence.

XX
KM Brain specific potassium channel; central nervous system disorder;

XX
XX dementia; cerebral ischemic sclerosis; therapy; probe; ss.

XX
OS Rattus sp.

XX
PN MO9937677-A1.

XX
PD 29-JUL-1999.

XX
PF 20-JAN-1999; 99MO-JP00190.

XX
PR 04-DEC-1998; 98JP-0346198.

XX
PR 23-JAN-1998; 98JP-0011434.

XX
PA (YAMA) YAMANOCHI PHARM CO LTD.

XX
PI Miyake A, Mochizuki S, Yokoi H;

XX
DR WPI; 1999-458683/38.

PT Potassium channel protein expressed specifically in brain tissue and
PT method for its production

XX
XX
PS Example 3; Page 57-59; 63pp; English.

CC This sequence encodes a rat potassium channel protein, and was used as a
CC probe to isolate DNA encoding the protein of the invention. The protein
CC of the invention is a human brain specific potassium channel protein. The
CC protein is used to treat and investigate disorders of the central nervous
CC system such as dementia and cerebral ischemic sclerosis.

XX
SQ Sequence 3736 BP; 706 A; 1210 C; 1000 G; 820 T; 0 other;

Query Match 11.5%; Score 36.2; DB 20; Length 3736;
Best Local Similarity 48.8%; Pred. No. 0.47;
Matches 98; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 13 TGTGAGCCCATGCCATGTGCACACCCACTTGGCTCTCTTAGCCTACACCTGCC 72
DB 3132 TGTGAGCCCATGCCATGTGCACACCCACTTGGCTCTCTTAGCCTACACCTGCC 3191
QY 73 CTTTCTTGGCCCTTGTAGTTACTGTGCTGAGATGATGAGCCCACTCCCAAGCATGGC 132
DB 3192 CTGGCTTGGCCCAAGCCTGTCTGAGGCTTATTTACTTGCATCCAGAGAGAGC 3251
QY 133 CTGAGACATGTCTCTCTGCCCCCTCCACAGCAGTATGATGATGATGATGATG 192
DB 3252 TGGGCTCTTGGCCCTTGGCTTGGGCTGAGCAGCTGAGCTGCTGCTGCTGCTG 3311
QY 193 GCAGAGAGACTGTGTACAA 213
DB 3312 ATTCTCTGACTTTTAAACA 3332

RESULT 6

AAA10594/C
ID AAA10594 standard; DNA: 10732 BP.

XX
AC AAA10594;

XX
DT 29-JUN-2000 (first entry)

XX
DE Gene encoding a subunit of cellulose synthase.

XX
XX Cellulose synthase; cellulose production; increase yield; ds.

XX
OS Vigna angularis.

XX
PN JP2000060568-A.

XX
PD 29-FEB-2000.

XX
PF 26-AUG-1998; 98JP-0239998.

XX
PR 26-AUG-1998; 98JP-0239998.

XX
PA (MIZU) MIZUNO K.

XX
PA (OJIP) OJI PAPER CO.

XX
DR P-PSDB; AAY85179.

XX
DR WPI; 2000-342371/30.

XX
PS Claim 2; Page 14-21; 32pp; Japanese.

XX
XX This sequence represents a gene encoding a subunit of the cellulose
XX synthase complex of Vigna angularis. The invention relates to subunits of
XX cellulose synthetic equipment, that can be used to increase the amount of
XX cellulose synthesised by a plant. The proteins and genes encoding them
XX can also be used to improve the properties of the cellulose being
XX produced by a plant.

XX SQ Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;
Query Match 11.5%; Score 36.2; DB 21; Length 10732;
Best Local Similarity 26.9%; Pred. No. 0.69;
Matches 39; Conservative 50; Mismatches 56; Indels 0; Gaps 0;
OY 20 CCATGCGATGACACCACTTGCCCTTCCTTGAAGCCCTGACCCCTTCTT 79
DB 9397 BYSTGCBYNSITDAYSRTYCTYCTCDYSSCTYTCACACTBCNSTSRKSKRTT 9338
OY 80 TGCCCTTGTAGTACTGTGCTGAGATGATGAGCGCCACTCCANAGCATGCGCTGAAC 139
DB 9337 YRCYCYTSRCYTCYTCYTCNCSRCCSRTTSRSRYADYSCNMCNCSRDYSPBCTTYDAC 9278
OY 140 AATGTCCCTCGCCCTCCACACAG 164
DB 9277 YTCYTYSDCYTCNCSCTCYTCSCSR 9253
RESULT 7
ID AAK80111 standard; DNA; 9545 BP.
XX AAK80111;
XX 07-NOV-2001 (first entry)
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34933.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytosolic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
XX MO200157182-A2.
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.

PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM,

DR WPI; 2001-483426/52.

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -

PS Disclosure; SEQ ID NO 34923; 3071pp + Sequence Listing; English.

CC AAK5495 to AAK64702 encode the human immune/hematopoietic antigen (I
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat Immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 9545 BP; 2568 A; 2101 C; 2253 G; 2623 T; 0 other;
XQ

Query Match	10.58; Score 33; DB 22; Length 9545;
-------------	--------------------------------------

Matches 42; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 36 CCCACTTTGGCCCTTTCCTTAGGCGCTACACCCTGCCCTTTCCTTGCCGCCTGTAGT 92
 ||||| | ||||| ||||| | ||||| | ||||| |
Db 1354 CCCACTAAGTAAGTCCCCCTTCGGCGTAGAACCTGCCCTTTCCTTGTATAT 1410

RESULT 8
AAI99683

ID	AAI99683	standard; DNA; 4403765 BP.
XX		
AC	AAI99683;	

DT 15-JAN-2002 (first entry)

DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2

KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds

05 Mycobacterium tuberculosis.

PN US6294328-B1

PD 25-SEP-2001.

PF 24-JUN-1998; 98US-0103840.

PR 24-JUN-1998; 98US-0103840.
VY

PA (GENO-) INST GENOMIC RES.
.YY

PI Fleischmann RD, White OR, Fraser CM, Venter JC;
YY

DR WPI; 2001-647261/74.
XX

PT Evaluating strain variation of *Mycobacterium tuberculosis*, comprises determining the nucleotide sequence of the strain at positions in the

PT genome corresponding to positions where *M. tuberculosis* strains CDC 1551 and H37Rv differ -

PS Claim 4: SEO TD NO 2: 3nn + Sequence Listing: English
XX

XX The invention relates to evaluating strain variation within and between

different populations of the tuberculosis bacterial pathogen, *Mycobacterium tuberculosis* or related *Mycobacterium* by determining the

CC nucleotide sequence of the first strain in the complete
CC sequence of the genome that correspond to positions that differ in the
CC

CC nucleotide sequences of *M. tuberculosis* strains CDC 1551 (A199683) and
CC H37Rv (A199682). The method is useful for evaluating strain variation of
CC

CC M. tuberculosis and has valuable application in the fields of

CC tuberculosis genetics, epidemiology, patient treatment and epidemic

CC

CC Note: The sequence data for this patent did not form part of the printed
CC monitoring.

at seqdata.uspto.gov/sequence.html?DocID=62294328B1.

Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other; SQ

Query match 10.5%; Score 33; DB 22; Length 4403765;

```

Best Local Similarity  22.08;  Freq: NO: 34;
Matches  63;  Conservative  0;  Mismatches  50;  Indels  0;  Gaps  0

```

33 ACACCCACTTGGCCTCTTCCCTTAGGCCCTACACCCCTGCCCTTCTTTGCCCTTGAGT 92

Db 1838235 ACGCCGACGAGCGTCTTCGTTGCCGCCGACACCACTGACGATGGTTTGGCGTCTTGTTGC 183

93 TACTGTGGCTGGAGTGATGAGCGCCACTCCCAAGCATGGCCTGGAACAATGTC 145

Db 1838295 CAGCGTTCGGGTAGAAATCACCGGCGATGTCGACAACGCCCTAAACAGTGTCTC 1838347

RESUL 3
AAI99682

XX ID AM199082 SCANDIAU; DNA; 4411029 BF.

AA15500Z;
AC
XX

DI 15-JAN-2002 (first entry)
XX

PN		WO200233097-A2.
XX		
XX		25-APR-2002.
PD		
PF		16-OCT-2001; 2001MO-EPI1926.
XX		
PR		16-OCT-2000; 2000US-240074P.
XX		13-NOV-2000; 2000US-247004P.
PA	(FARB) BAYER AG.	
XX		
PI	Zhu Z;	
DRI	WPI; 2002-435539/46.	
PT	New human sulfotransferase polypeptides and polynucleotides, useful for treating disorders such as cancer and chronic obstructive pulmonary disorder, and in screening for (ant)agonists of the enzyme -	
PS	Disclosure; Fig 6; 118pp; English.	
XX	The invention describes an isolated polynucleotide encoding a sulfotransferase polypeptide. Human sulfotransferase is useful for treating disorders such as cancer, a central nervous system (CNS) disorder, diabetes, asthma, and chronic obstructive pulmonary disease (COPD), and in screening for sulfotransferase agonists and antagonists. The proteins can be used to immunise a mammal to produce polyclonal antibodies, and in diagnostic assays for detecting diseases and abnormalities or susceptibility to diseases and abnormalities related to the presence of mutations in the nucleic acids encoding the enzymes. The expression vector comprising a DNA encoding sulfotransferase, and the reagent which modulates the activity of sulfotransferase are useful in preparing a medicament for modulating the activity of a sulfotransferase in a disease, such as cancer, and COPD. This sequence encodes a fragment of the human sulfotransferase studied in the invention.	
SO	Sequence 378 BP; 85 A; 136 C; 89 G; 68 T; 0 other:	
	Query Match	10.3%; Score 32.6; DB 24; Length 378;
	Best Local Similarity	52.6%; Pred. No. 2.8;
	Matches 71; Conservative	0; Mismatches 64; Indels 0; Gaps 0
OY	33 ACACCCACTTGGCGCTTCCTCAAGGCGCTCACACCCTGCCCTTTTGCCCTGTGAT	92
DB	143 ACCCTTAATTCTAACAACATCGTCACGCGCTACTCCAACGGCTTCGCCTCCGAAG	202
OY	93 TACTGTGGCTGGAGTAGTAGAGCGCATCTCCCAAGCATGGCCTCGAACATGTCTCTGCG	152
DB	203 CCTCGGCCACAGGCGCTTCTGGGGCCACCTGGGCGACGCCACAGGAAAGCACTTCGCCCTGCG	262
OY	153 CCCTCCACAGCAGT 167	
DB	263 GCCTCCCTGGCGCACT 277	
RESULT 11		
ABA20784/C		
ID	ABA20784 standard; DNA; 2633 .BP.	
XX		
DE	Human nervous system related polynucleotide SEQ ID NO 13115.	
XX		
KM	Human; nootropic; neuroprotective; cytosstatic; dermatological; virocidic; immunosuppressive; antiinflammatory; anti-RHV; antibacterial; vulnerary; antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer; antifungal; hepatoprotective; cerebroprotective; antiinflammatory; antiallergic; antididiabetic; anticancer; anticoagulant; antifungal; antiparasitic; cardiant; immune disorder; cardiovascular disorder;	

KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
 XX Homo sapiens.
 XX WO200159063-A2.
 PD 16-AUG-2001.
 XX 17-JAN-2001; 2001WO-US01334.
 PF
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209457.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 20-OCT-2000; 2000US-0242221.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 03-JAN-2001; 2001US-0259678.

[illegible]

XX			16-OCT-2000; 2000US-240074P.
PR			13-NOV-2000; 2000US-247004P.
XX			(FARB) BAYER AG.
PA			Zhu Z;
XX			WIPI: 2002-435539/46.
DR			P-PsDB; AAU96788.
XX			New human sulfotransferase polypeptides and polynucleotides, useful for
PT			treating disorders such as cancer and chronic obstructive pulmonary
PI			disorder, and in screening for (anti)agonists of the enzyme
PS			Claim 1; Fig 13; 118pp; English.
XX			The invention describes an isolated polynucleotide encoding a
CC			sulfotransferase polypeptide. Human sulfotransferase is useful for
CC			treating disorders such as cancer, a central nervous system (CNS)
CC			disorder, diabetes, asthma, and chronic obstructive pulmonary disease
CC			(COPD), and in screening for sulfotransferase agonists and antagonists.
CC			The proteins can be used to immunise a mammal to produce polyclonal
CC			antibodies, and in diagnostic assays for detecting diseases and
CC			abnormalities or susceptibility to diseases and abnormalities related
CC			to the presence of mutations in the nucleic acids encoding the enzymes.
CC			The expression vector comprising a DNA encoding sulfotransferase, and
CC			the reagent which modulates the activity of sulfotransferase are useful
CC			in preparing a medicament for modulating the activity of a
CC			sulfotransferase in a disease, such as cancer, and COPD. This sequence
CC			encodes the human sulfotransferase studied in the invention.
XX			Sequence 4712 BP; 1120 A; 1225 C; 1120 G; 1247 T; 0 other;
SQ			
	Query Match	10.3%; Score 32.6; DB 24; Length 4712;	
	Best Local Similarity	52.6%; Pred. No. 6.9; Mismatches 64; Indels 0; Gaps 0;	
	Matches	71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;	
OY		33 ACACCCACTGCGCTTCCTTGAAGGCCTACACCCTGCCCTTTTGCCCTGTGAGT	92
Dd		1213 ACCCTTAACCTCAACAACATCATCGGCTCTACTCCAACGGCTCCGCGCACCTCGACG	1272
OY		93 TAGTGTGGCTGGAGATGAGACGGCCACATCCCAGAAGTAGGCTGGAAAGATGCTGCCTGNC	152
Dd		1273 CCTGTGCGAAGAGCCCTTCTGGGGGCCACTGGCGCAGCGCAGGAGACATTCCGCTGC	1332
OY		153 CCCCTCACCAAGCAGT	167
Dd		1333 GCTGCTGCCCCGACCT	1347
RESULT 15			
AAS94823			
ID	AAS94823 standard; DNA; 4748 BP.		
XX	AAS94823;		
AC			
XX	14-FEB-2002 (first entry)		
DT			
XX			
DE	Human DNA sequence #78 expressed during foam cell differentiation.		
XX			
KW	Human; foam cell differentiation; atherosclerosis; cerebral stroke;		
RK	cardiovascular disorder; coronary artery disease; gene therapy; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	MO200177389-A2.		
PD	18-OCT-2001.		
PF	04-APR-2001; 2001WO-US11128.		
PR	05-APR-2000; 2000US-195106P.		

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:25:41 ; Search time 22 seconds

(without alignments)
197.955 Million cell updates/sec

Title: US-09-955-807-2

Perfect score: 582

Sequence: 1 MLGSEMPKPCNPLGLFLG.....PHPWLTWGHILRNHYSSASH 105

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	80	13.7	264	1	SNAIL_MOUSE
2	77	13.2	388	1	SOX7_HUMAN
3	76	13.1	380	1	SOX7_MOUSE
4	73.5	12.6	323	1	OLG2_HUMAN
5	72	12.4	575	1	MIS_PIG
6	71.5	12.3	989	1	DLR4_HUMAN
7	71.5	12.3	992	1	DLR4_RAT
8	71	12.2	1964	1	NTC4_MOUSE
9	69	11.9	762	1	E13B_TRIHA
10	69	11.9	1032	1	S24D_HUMAN
11	68.5	11.8	323	1	OLG2_MOUSE
12	67	11.5	357	1	MDL1_ARATH
13	67	11.5	2109	1	PCCA_CHICK
14	66.5	11.4	1841	1	CC12_SCHPO
15	66	11.3	368	1	GP62_HUMAN
16	66	11.3	686	1	VGIL_PVRN3
17	66	11.3	829	1	CADG_HUMAN
18	66	11.3	913	1	PCK5_HUMAN
19	66	11.3	1877	1	PCK5_MOUSE
20	66	11.3	1877	1	HRX_MOUSE
21	66	11.3	3865	1	DMD_HUMAN
22	65.5	11.3	3685	1	DMD_HUMAN
23	65	11.2	686	1	VGIL_PVRN3
24	65	11.2	1324	1	IRS2_HUMAN
25	65	11.2	2239	1	CCAE_RABIT
26	64.5	11.1	267	1	THIG_AQUAE
27	64.5	11.1	380	1	DAPA_MAIZE
28	64.5	11.1	875	1	NERF_HUMAN
29	64.5	11.1	1077	1	AT10_HUMAN
30	64	11.0	380	1	VASP_HUMAN
31	64	11.0	445	1	GUDH_ECOLI
32	64	11.0	607	1	HEMA_CDVA4
33	63.5	10.9	245	1	HXB4_CHICK

34	63.5	10.9	384	1	VASP_CANFA	P50551 canis fam1
35	63.5	10.9	453	1	ARX_BRARE	O42115 brachydant
36	63.5	10.9	456	1	KICH_HUMAN	P35790 homo sapien
37	63.5	10.9	1447	1	PTCL_HUMAN	Q13635 homo sapien
38	63	10.8	453	1	KICH_RAT	Q01134 ratius norv
39	63	10.8	1434	1	PTCL_MOUSE	O61115 mus musculu
40	63	10.8	1736	1	CAZB_HUMAN	P13942 homo sapien
41	63	10.8	3678	1	DMD_HUMAN	P11531 mus musculu
42	62.5	10.7	116	1	BRCL_HUMAN	O99218 homo sapien
43	62.5	10.7	346	1	HM22_CAEL	P41936 caenorhabdi
44	62.5	10.7	372	1	3BR3_RAT	P27364 ratius norv
45	62.5	10.7	443	1	FD6C_BRANA	P48627 brassica na

ALIGNMENTS

RESULT 1

ID	SNAIL_MOUSE	STANDARD:	PRT:	264 AA.
AC	002085;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	zinc finger protein SNAIL (Snail protein homolog) (Sna protein).			
GN	SNAIL OR SNA			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid:10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=embryo;			
RX	MEDLINE-93201990; PubMed-1295727;			
RA	Smith D.E., del Amo F.F., Gridley T.;			
RT	"Isolation of Sna, a mouse gene homologous to the Drosophila genes			
RT	snail and escargot: its expression pattern suggests multiple roles			
RT	during postimplantation development.";			
RL	Development 116:1033-1039(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/Sv;			
RX	MEDLINE-97419272; PubMed-9271672;			
RA	Nieto A.M., Bennett M.F., Sargent M.G., Wilkinson D.G.;			
RT	"Cloning and developmental expression of Sna, a murine homologue of			
RT	the Drosophila snail gene.";			
RL	Development 116:227-237(1992).			
RN	[3]			
RP	SEQUENCE OF 1-27 FROM N.A.			
RC	STRAIN=129/Sv;			
RX	MEDLINE-97419272; PubMed-9271672;			
RA	Jiang R., Copeland N.G., Gilbert D.J., Jenkins N.A., Gridley T.;			
RT	"Genomic organization and chromosomal localization of the mouse snail			
RT	(Sna) gene.";			
RL	Mamm. Genome 8:686-688(1997).			
CC	-1- FUNCTION: THIS PROTEIN HAS MANY ROLES DURING POSTIMPLANTATION			
CC	DEVELOPMENT. IT IS INVOLVED IN EMBRYONIC MESODERM FORMATION			
CC	AND ITS MAINTENANCE AND MAY ALSO BE INVOLVED IN CHONDROGENESIS			
CC	AND IN EPITHELIAL-MESENCHYMAL INDUCTIVE INTERACTIONS.			
CC	-1- SUBCELLULAR LOCATION: Nuclear (potential).			
CC	-1- SIMILARITY: BELONGS TO THE SNAIL FAMILY OF ZINC FINGER PROTEINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
DR	EMBL; M55604; AAA03481.1; -			
DR	EMBL; X67253; CAA47675.1; -			
DR	EMBL; U95961; AAB58054.1; -			

DR TRANSFAC; T02333; -
 DR MGD; MGI:98330; Sna.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf_C2H2_4.
 DR PRINTS; PR00046; ZINC_FINGER.
 DR PRODOM; PD000003; Znf_C2H2_1.
 DR SMART; SM00355; Znf_C2H2; 3.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
 DR Developmental protein; Zinc-finger; Metal-binding; DNA-binding;
 KM Nuclear protein; Repeat.
 FT DOMAIN 156 259 ZINC_FINGERS.
 FT ZN_FING 156 176 C2H2-TYPE.
 FT ZN_FING 180 202 C2H2-TYPE.
 FT ZN_FING 210 230 C2H2-TYPE.
 FT ZN_FING 236 259 C2H2-TYPE (ATYPICAL).
 FT CONFLICT 219 219 D -> V (IN REF. 2).
 SQ SEQUENCE 264 AA; 29190 MW; 52E2061224A18DB CRC64;

Query Match 13.7%; Score 80; DB 1; Length 264;
 Best Local Similarity 24.5%; Pred. No. 0.76;
 Matches 46; Conservative 13; Mismatches 37; Indels 92; Gaps 11;

QY 4 YSPMCAHPLGLGFLG-----LHPALSLP-----LVTVYAGVMSAT-----PK- 42
 DB 31 FQDPYDQAH-----LAAIPPEVLNPNASLPTLIMDSLLVPOVPVAMATPLRESPKA 85
 QY 43 -----HGLEQCPAPP-PAYTGFTGDSG-----AKETVS 70
 DB 86 VELTSLSDSDSGSKSQSP 145
 QY 71 QDKRS-----QGH-----TCTCLALPHPLWTVGHLRNHVS 101
 DB 146 KDPOSCKRIENCKYCNKEVLSGLAKMHIRSHLTPLCVCTGKAFSRPML-QGHVFRHTG 204
 QY 102 -----SASH 105
 DB 205 EKPFSCSH 212

RESULT 2

SOX7_HUMAN STANDARD; PRT; 388 AA.
 ID SOX7_HUMAN
 AC Q9BT81;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription factor SOX-7.
 GN SOX7.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thymus;
 RA Tashiro H., Yamazaki Y., Watanabe K., Kumagai A., Itakura S.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Nagatsuma M.,
 RA Kawakami B., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
 RA Isogai T., Suganuma S., Nagahara K., Masuno Y., Nagai K.,
 RA Isogai T.,

RT "MEDO human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RA Strausberg R.;

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Transcriptional repressor. Binds to the DNA sequence 5'-

CC AACCAT-3'.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: CONTAINS 1 HMG BOX.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AJ409320; CAC84226.1; -
 CC EMBL; AK055556; BAB70955.1; -
 CC EMBL; BC004299; AAH04299.1; -
 CC HSP; P48436; ISX9.
 CC Genew; HGNC:18196; SOX7.
 CC InterPro: IPR000910; HMG_12_Box.
 CC Pfam; PF00505; HMG_Box; 1.
 CC SMART; SM00398; HMG; 1.
 CC DNA-binding; Nuclear protein; Transcription regulation; Repressor.
 FT DNA_BIND 45 113 HMG_BOX.
 SQ SEQUENCE 388 AA; 42197 MW; DBADEFB440DC2A74 CRC64;

Query Match 13.2%; Score 77; DB 1; Length 388;
 Best Local Similarity 37.0%; Pred. No. 2.2;
 Matches 17; Conservative 5; Mismatches 10; Indels 14; Gaps 2;

QY 7 PMCAHPLGLGLPLSLPLVTVYAGVMSATPKHGOCPAP 52
 DB 250 PLHCSPHPLGLSLALGOSPGVSM-----MSPPV-----GCPSP 281

RESULT 3

SOX7_MOUSE STANDARD; PRT; 380 AA.
 ID SOX7_MOUSE
 AC P40646; Q9R1T6;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription factor SOX-7 (mSOX7).
 GN SOX7 OR SOX-7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=99255428; PubMed=10320775;
 RA Taniguchi K., Hirooka Y., Ogawa M., Sakai Y., Kido S., Aiso S.;
 RT "Isolation and characterization of a mouse SRV-related cDNA, mSOX7.";
 RL Biochim. Biophys. Acta 1445:225-231(1999).
 RN [2]
 RP SEQUENCE OF 56-109 FROM N.A.
 RC STRAIN=Parkes; TISSUE=Brain, and Testis;
 RX MEDLINE=92310993; PubMed=1614875;
 RA Denny P., Swift S., Brand N., Dabhadre N., Barton P., Ashworth A.;
 RT "A conserved family of genes related to the testis determining gene,
 RT Nucleic Acids Res. 20:2887-2887(1992).
 RL -1- FUNCTION: Transcriptional repressor. Binds to the DNA sequence 5'-
 CC AACCAT-3'.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: Ovary and heart.
 CC -1- SIMILARITY: CONTAINS 1 HMG_BOX.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB023419; BAA78765.1; -
DR EMBL; X65660; CAA46611.1; -
DR PIR; S21487; S21487.
DR PIR; S22945; S22945.
DR HSSP; P48436; 1S9M.
DR MGD; MGI:98369; Sox7.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_Box; 1.
DR SMART; SM00398; HMG; 1.
DR DNA-BINDing; Nuclear protein; Transcription regulation; Repressor.
FT DNA_BIND 45 113 HMG_BOX.
SQ SEQUENCE 380 AA; 41489 MW; BE780D7234CB895C CRC64;

Query Match 13.1%; Score 76; DB 1; Length 380;
Best Local Similarity 24.0%; Pred. No. 2.7;
Matches 31; Conservative 9; Mismatches 31; Indels 58; Gaps 6;

OY 7 PMCAHPGLFLGLHPALSLPLVTVAGVMSATPKHGLEOCPE----- 50
DB 242 PLKSHPLGSLALGOSPEVS--MMSVSG-----CPSPAYSHATYPLHPN 287
OY 51 -----APPAYVTF-----TGDSGAKETVSODKRSIGHTWCTLALPHPWLT 91
DB 288 LQAHLGQSPPEPHGFDTLDQLSQVELLGDMDRNE-PDQYLNKPGHDSAAAG----- 339
OY 92 WVGHRNRY 100
DB 340 -VGTLTGHV 347

RESULT 4
OLG2_HUMAN STANDARD; PRT; 323 AA.
ID OLG2_HUMAN Q13516; Q9N214;
AC 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Oligodendrocyte lineage transcription factor 2 (Basic helix-loop-helix
DE protein class B 1) (Protein kinase C-binding protein RACK17) (Protein
DE kinase C binding protein 2).
GN OLG2 OR BHLH1 OR PRKCBP2 OR RACK17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Hippocampus;
RA Kuroda S., Tokunaga C., Kiyohara Y., Konishi H., Kikawa U.;
RT "Protein kinase C-binding protein.";
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20202664; PubMed=10737801;
RA Wang J., Jani-Sait S.N., Escalon E.A., Carroll A.J., de Jong P.J.,
RA Kirsch I.R., Apian P.D.;
RT "The t(14;21)(q11.2;q22) chromosomal translocation associated with
RT T-cell acute lymphoblastic leukemia activates the BHLH1 gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:3497-3502(2000).
CC -1- FUNCTION: Required for oligodendrocyte and motor neuron
CC specification in the spinal cord (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- DISEASE: A form of T-cell acute lymphoblastic leukemia (T-ALL) is
CC characterized by a chromosomal translocation t(14;21)(q11.2;q22)

```

```

CC -----
CC which involves OLG2 and T-cell receptor alpha chain (TCRA) genes.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48250; AAC72247.1; ALT_INIT.
DR EMBL; AF221520; AAF61215.1; ALT_INIT.
DR Genew; HGNC:9398; OLG2.
DR MIM; 606386; -
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS50888; HLH_2; 1.
DR Transcription regulation; DNA-binding; Nuclear protein;
DR Proto-oncogene; Chromosomal translocation.
FT DNA_BIND 109 120 BASIC DOMAIN.
FT DOMAIN 121 162 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 77 88 POLY-SER.
FT DOMAIN 206 212 POLY-ALA.
FT DOMAIN 225 237 POLY-ALA.
FT DOMAIN 266 270 POLY-ALA.
FT DOMAIN 273 280 POLY-GLY.
SQ SEQUENCE 323 AA; 32384 MW; 0EC9223961062509 CRC64;

Query Match 12.6%; Score 73.5; DB 1; Length 323;
Best Local Similarity 25.5%; Pred. No. 4;
Matches 35; Conservative 12; Mismatches 35; Indels 55; Gaps 7;

OY 2 LGYSEPMPC--AHPLGLFLGLHPALSLPLV-----VTVAGVMSAT----- 40
DB 192 LAHSAPLPAAATRAAANAHAHHPAVHPIPLPPAAAAAASASIPSGSLPSV 251
OY 41 ----PKHGLEOCPPAPPPA-VTFTGDSGAKETVSODKRSOGH-----TWCTLAL 85
DB 252 GSIRPPHGLIKSPSAAAAPIAGCGGCGSGA-----SGGFQHGMPGPCSCMCQVPP 302
OY 86 PHPWLTWVGHLRNHYSS 102
DB 303 PH-----HHVSA 309

RESULT 5
MIS_PIG STANDARD; PRT; 575 AA.
ID MIS_PIG P79295;
AC 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Muellerian inhibiting factor precursor (MIF) (Anti-muellerian hormone)
DE (AMH) (Muellerian inhibiting substance).
GN AMH.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Daneau I., Silversides D.W.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS GLYCOPROTEIN, PRODUCED BY THE SEROTIN CELLS OF THE
CC TESTIS, CAUSES REGRESSION OF THE MUELLERIAN DUCT. IT ALSO IS ABLE
CC TO INHIBIT THE GROWTH OF TUMORS DERIVED FROM TISSUES OF MUELLERIAN
CC DUCT ORIGIN (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U08053; AAC25968.1; -
DR InterPro: IPR001839; TGFb.
DR Pfam: PF000019; TGF-beta; 1.
DR ProDom: PD000357; TGFb; 1.
DR SMART: SM00204; TGFb; 1.
DR PROSITE: PS00250; TGF_BETA_1; 1.
DR Growth factor: Glycoprotein; Gonadal differentiation; signal.
KW SIGNAL
FT PROPEP 17 23
FT CHAIN 24 525
FT DISULFID 477 541
FT DISULFID 503 572
FT DISULFID 507 574
FT DISULFID 540 540
FT CARBOHYD 78 78
FT CARBOHYD 343 343
SQ SEQUENCE 575 AA; 61504 MW; 69AF6354B390780 CRC64;

Query Match
Best Local Similarity 29.1%; Score 72; DB 1; Length 575;
Matches 32; Conservative 9; Mismatches 33; Indels 36; Gaps 5;

QY 1 MLCYSEMPMC-ANPLFLGLHPALSLPLVTVAGVMSATPRHGLECPAPPAVATGTF 59
DB 159 LKFGELPLPEASPLALVLYPRGPETVYAGLPGAG-----SLCC----- 203
QY 60 TGDSGAKETVYSDKRSQGHWTCTALPHPLVLTWVGH-----LRNHVSAS 104
DB 204 TRNSG-----FLALAVDRPRANRGSLATLRKNGAS 238

RESULT 6
DLP4_HUMAN STANDARD; PRT; 989 AA.
AC G9Y2H0; G9H1L7; G9H137; G9H138;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Disks large-associated protein 4 (DAP-4) (SAP90/PSD-95-associated
DE protein 4) (SAPAP4) (PSD-95/SAP90 binding protein 4).
GN DLAGP4 OR DAP4 OR KIA00964.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ON NCBI_TaxID=9606;
RX TISSUE=Brain;
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,

```

```

RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Laylor S.,
RA Leharalho M.H., Leverisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McIlroy K., McKay A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilmshurst L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
CC -1- FUNCTION: May play a role in the molecular organization of
CC synapses and neuronal cell signaling. Could be an adapter protein
CC linking ion channel to the subnaptic cytoskeleton. May induce
CC enrichment of PSD-95/SAP90 at the plasma membrane.
CC -1- SUBUNIT: Interacts with DLG1 and DLAG4/PSD-95 (by similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-associated (by similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE SAPAP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB023181; BAA76808.1; -
DR EMBL; AL390374; CAC18627.1; -
DR EMBL; AL050318; CAB75370.1; -
DR EMBL; AL050318; CAC16960.1; -
DR InterPro: IPR005026; GKAP.
DR Pfam: PF03359; GKAP; 1.
KW Membrane; Alternative splicing.
FT DOMAIN 267 274
FT VARSPIC 671 697
FT CONFLICT 229 229
FT SEQUENCE 989 AA; 107578 MW; B391716801FDEC1B CRC64;

Query Match
Best Local Similarity 31.6%; Score 71.5; DB 1; Length 989;
Matches 25; Conservative 6; Mismatches 39; Indels 9; Gaps 3;

QY 15 GLFLGLHPALSLPLVTVAGVMSATPRHG-LECCPPAPPAVATGTFDGSAGKET 68
DB 226 GLMILGROARSPQRFEMHAYNTISCHMLKTTNNTELTAPPPAPAPATCPSLGVGD 285
QY 69 VSODKRSQGHWTCTALPH 87
DB 286 TNYVKG---SMSTLTLSH 301

RESULT 7
DLP4_RAT STANDARD; PRT; 992 AA.
AC P97839;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Disks large-associated protein 4 (DAP-4) (SAP90/PSD-95-associated

```


an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(Tw) and a N-terminal fragment N(ETC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane.

-1- PSM: Phosphorylated.

-1- DISEASE: loss of the extracellular domain causes constitutive activation of the notch protein, which leads to hyperproliferation of glandular epithelial tissues and development of mammary carcinomas.

-1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.

-1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.

-1- SIMILARITY: CONTAINS 3 LIN/NORCH REPEATS.

-1- SIMILARITY: CONTAINS 5 ANK REPEATS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M80456; AAB38377.1; -

EMBL; U43691; AAC52630.1; -

EMBL; U43691; AAC52631.1; -

EMBL; AF030001; AAB82004.1; -

EMBL; AB016771; BAA32281.1; -

EMBL; AB016772; BAA32283.1; -

EMBL; AB016773; BAA32284.1; -

EMBL; AB016774; BAA32285.1; -

EMBL; A38072; TBMV73.

HSP: P08709; 1BF9.

MCD; MGI:107471; Notch4.

InterPro: IPR002110; ANK.

InterPro: IPR000152; Asx_hydroxyl.

InterPro: IPR000561; EGF-Like.

InterPro: IPR000742; EGF_2.

InterPro: IPR001881; EGF_Ca.

InterPro: IPR001438; EGF_L1.

InterPro: IPR000800; Notch.

Pfam: PF00008; EGF; 27.

Pfam: PF00066; ank; 6.

PRINTS: PR01415; ANKYRN.

PRINTS: PR00010; EGFBL00.

PRINTS: PR01452; NOTCH.

SMART; SM00248; ANK; 5.

SMART; SM00179; EGF_CA; 11.

SMART; SM00001; EGF_Like; 15.

SMART; SM00004; NL; 2.

PROSITE; PSS0088; ANK_REPEAT; 5.

PROSITE; PSS0297; ANK_REPEAT_REGION; 1.

PROSITE; PS00010; ASX_HYDROXYL; 11.

PROSITE; PS00022; EGF_1; 28.

PROSITE; PS01166; EGF_2; 21.

PROSITE; PS01187; EGF_CA; 9.

Receptor: Transcription regulation; Activator; Differentiation; Developmental protein; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.

CHAIN 1 20

CHAIN 21 1964

CHAIN 1411 1964

CHAIN 1428 1964

CHAIN 1463 1964

CHAIN 21 1443

CHAIN 1444 1464

CHAIN 1465 1964

CHAIN 21 60

CHAIN 61 112

EGF-Like 1.

EGF-Like 2.

FT DOMAIN 115 152 EGF-Like 3.

FT DOMAIN 153 189 EGF-Like 4.

FT DOMAIN 191 229 EGF-Like 5.

FT DOMAIN 231 271 EGF-Like 6.

FT DOMAIN 273 309 EGF-Like 7.

FT DOMAIN 311 350 EGF-Like 8.

FT DOMAIN 352 388 EGF-Like 9.

FT DOMAIN 389 427 EGF-Like 10.

FT DOMAIN 429 470 EGF-Like 11.

FT DOMAIN 472 508 EGF-Like 12.

FT DOMAIN 510 546 EGF-Like 13.

FT DOMAIN 548 584 EGF-Like 14.

FT DOMAIN 586 622 EGF-Like 15.

FT DOMAIN 623 656 EGF-Like 16.

FT DOMAIN 658 696 EGF-Like 17.

FT DOMAIN 724 762 EGF-Like 18.

FT DOMAIN 726 762 EGF-Like 19.

FT DOMAIN 764 800 EGF-Like 20.

FT DOMAIN 803 839 EGF-Like 21.

FT DOMAIN 841 877 EGF-Like 22.

FT DOMAIN 878 924 EGF-Like 23.

FT DOMAIN 926 962 EGF-Like 24.

FT DOMAIN 964 1000 EGF-Like 25.

FT DOMAIN 1002 1040 EGF-Like 26.

FT DOMAIN 1042 1081 EGF-Like 27.

FT DOMAIN 1083 1122 EGF-Like 28.

FT DOMAIN 1126 1167 EGF-Like 29.

FT REPEAT 1168 1208 LIN/NORCH 1.

FT REPEAT 1209 1242 LIN/NORCH 2.

FT REPEAT 1243 1282 LIN/NORCH 3.

FT REPEAT 1628 1657 ANK 1.

FT REPEAT 1661 1691 ANK 2.

FT REPEAT 1695 1724 ANK 3.

FT REPEAT 1728 1757 ANK 4.

Query Match 12.28; Score 71; DB 1; Length 1964;

Best Local Similarity 24.88; Pred. NO. 45;

Matches 25; Conservative 4; Mismatches 36; Indels 36; Gaps 5;

QY 3 GYSEPMPCAHPLGLFLGLHPALSLPLVTVAGVMSATPKHGLE---GCPAPP----- 53

Db 1129 GCGPSPCLH-----NCGTCTPELGNGPGCTCPDPSGPRCQ 1167

QY 54 -PAVTGFTGDSGAKETYSQKRSOGHTW---CTLALPHW 89

Db 1168 RFGASGCEG-RGGDGTCDACSGPGGDMGDCSLGVDPDW 1207

RESULT 9

ID E13B_TRIHA STANDARD; PRT; 762 AA.

AC P53626;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Glucan endo-1,3-beta-glucosidase BGN13.1 precursor (EC 3.2.1.39)

DE ((1->3)-beta-glucan endohydrolase BGN13.1) ((1->3)-beta-glucanase

DE BGN13.1) (Basic beta-1,3-endoglucanase BGN13.1).

GN BGN13.1.

OS Trichoderma harzianum.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Hypocreales; mitosporic Hypocreales; Trichoderma.

OX NCBI_Taxid=5544;

RN [1]

RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RC STRAIN=CECT 2413;

RA MEDLINE=96074325; PubMed=7592488;

RA de la Cruz J., Pincor-Toro J.A., Benitez T., Llobell A., Romero L.C.;

RT "A novel endo-beta-1,3-glucanase, BGN13.1, involved in the

RT mycoparasitism of Trichoderma harzianum.";

RL J. Bacteriol. 177:6937-6945(1995).

CC -1- FUNCTION: INVOLVED IN MYCOPARASITISM, HYDROLYZES YEAST AND FUNGAL

CC CELL WALLS. CLASSIFIED AS A SMALL-OLIGOSACCHARIDE-PRODUCING TYPE


```
CC BASED ITS THE END PRODUCTS: GLUCOSE, LAMINARIBIOSE OR
CC LAMINARITROSE. OPTIMUM TEMPERATURE FOR ACTIVITY IS 40 DEGREES
CC CELSIUS. AND INHIBITION OCCURS AT 55 DEGREES CELSIUS.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,3-beta-D-glucosidic linkages
CC in 1,3-beta-D-glucans.
CC -1- ENZYME REGULATION: INHIBITED BY GLUCOSE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE C-TERMINAL CYSTEINE-RICH REGION MAY FUNCTION AS A
CC FUNGAL CELL WALL BINDING DOMAIN.
CC -1- PFM: DOES NOT SEEM TO BE GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO FAMILY 55 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: X84085; CAA58889.1; -
CC KM Hydrolase; Glycosidase; Signal.
CC FT SIGNAL 1 16 POTENTIAL.
CC FT PROPEP 17 33
CC FT CHAIN 34 762 GLUCAN ENDO-1,3-BETA-GLUCOSIDASE BGN13.1.
CC FT DOMAIN 520 762 CYS-RICH.
CC FT CONFLICT 40 40 P -> F (IN REF. 1; AA SEQUENCE).
CC FT CONFLICT 759 759 R -> P (IN REF. 1; AA SEQUENCE).
CC SQ SEQUENCE 762 AA; 81246 MW; 0CC25C3C10897AFL CRC64;

Query Match 11.9%; Score 69; DB 1; Length 762;
Best Local Similarity 32.9%; Pred. No. 26;
Matches 24; Conservative 8; Mismatches 29; Indels 12; Gaps 3;

QY 10 CAHPG-----LFLGLHPALSLPLVTVAGVMSATPRH---GLECCPPAPPVATGFTG 61
DB 538 CLPYIGSGHMWLYQLNLHNAAN-----VYVSLQAEFTNHOGANTQIIPAPVAVANGVTWG 653
QY 62 DSGAKETVSQDKR 74
DB 654 DDFSWCNGGDKR 666

RESULT 10
S24D_HUMAN STRAND: PRT; 1032 AA.
AC 094855:
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Protein transport protein Sec24D (SEC24-related protein D).
GN SEC24D OR KIAA0755.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RX Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RX Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=99262157; PubMed=10329445;
RX Tang B.L., Kausalya J., Low D.T.H., Lock M.L., Hong W.;
RT "A family of mammalian proteins homologous to yeast Sec24p.";
RL Biochem. Biophys. Res. Commun. 258:679-684(1999).
```

```
CC -1- FUNCTION: COMPONENT OF THE COP1 COAT, THAT COVERS ER-DERIVED
CC VESICLES INVOLVED IN TRANSPORT FROM THE ENDOPLASMIC RETICULUM TO
CC THE GOLGI APPARATUS. COP1 ACTS IN THE CYTOPLASM TO PROMOTE THE
CC TRANSPORT OF SECRETORY, PLASMA MEMBRANE, AND VACUOLAR PROTEINS
CC FROM THE ENDOPLASMIC RETICULUM TO THE GOLGI COMPLEX.
CC -1- SUBUNIT: COP1 IS COMPOSED OF AT LEAST FIVE PROTEINS: THE SEC23/24
CC COMPLEX, THE SEC13/31 COMPLEX AND SAK1.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND PERINUCLEAR (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: UNBIOGENOUSLY EXPRESSED, WITH HIGHER AMOUNTS IN
CC PLACENTA, PANCREAS, HEART AND LIVER.
CC -1- SIMILARITY: BELONGS TO THE SEC23/SEC24 FAMILY. SEC24 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AB018298; BAA34475.1; -
CC DR EMBL: AF130464; AAD28756.2; -
CC DR Genew; HGNC:10706; SEC24D.
CC KM Transport; Protein transport; Golgi stack; Endoplasmic reticulum;
CC Multigene family.
CC FT DOMAIN 363 388 ZINC FINGER-LIKE.
CC FT DOMAIN 10 293 PRO-RICH.
CC SQ SEQUENCE 1032 AA; 112999 MW; 1EAA80215EB97947 CRC64;

Query Match 11.9%; Score 69; DB 1; Length 1032;
Best Local Similarity 28.4%; Pred. No. 36;
Matches 21; Conservative 7; Mismatches 30; Indels 16; Gaps 2;

QY 4 YSEPMCAHPGLFLGLHPALSLPLVTVAGVMSATPRHGLECCPPAPPVATGFTGDS 63
DB 29 YGDPSTHATPTGM-----MKPAPGLATATRGW-LPPGPPPGPHQFGON 72
QY 64 GAKETVSQDKRSOG 77
DB 73 GAHATGHPPOPPG 86

RESULT 11
OLG2_MOUSE STRAND: PRT; 323 AA.
AC 09EQW6; Q9JUK4;
DT 15-JUN-2002 (Rel. 41; Created)
DT 15-JUN-2002 (Rel. 41; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE Oligodendrocyte lineage transcription factor 2.
GN OLG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=20544962; PubMed=11091082;
RX Takebayashi H., Yoshida S., Sugimori M., Kosako H., Komihama R.,
RX Nakafuku M., Nabeshima Y.;
RT "Dynamic expression of basic helix-loop-helix Olig family members:
RT Implication of Olig2 in neuron and oligodendrocyte differentiation and
RT identification of a new member, Olig3.";
RL Mech. Dev. 99:143-148(2000).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20182808; PubMed=10719889;
RX Zhou Q., Wang S., Anderson D.J.;
RT "Identification of a novel family of oligodendrocyte lineage-specific
RT basic helix-loop-helix transcription factors.";
RL Neuron 25:331-343(2000).
-1- FUNCTION: Required for oligodendrocyte and motor neuron
```

CC specification in the spinal cord.
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
 CC -1- TISSUE SPECIFICITY: Specifically expressed in zones of
 CC neuroepithelium from which oligodendrocyte precursors emerge, as
 CC well as in the precursors themselves.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB038697; BAB18907.1; -
 CC EMBL: AF232929; AAF61722.1; -
 CC MGI: 1355331; O1192.
 CC InterPro: IPR001092; HLH_basic.
 CC Pfam: PF00010; HLH_1.
 CC SMART: SM00353; HLH_1.
 CC DR PROSITE: PS00038; HLH_1; 1.
 CC DR PROSITE: PS50888; HLH_2; 1.
 CC KW Transcription regulation; DNA-binding; Nuclear protein.
 CC FT DNA_BIND 109 120 BASIC DOMAIN.
 CC FT DOMAIN 121 162 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 CC FT DOMAIN 77 88 POLY-SER.
 CC FT DOMAIN 206 212 POLY-ALA.
 CC FT DOMAIN 225 237 POLY-ALA.
 CC FT DOMAIN 266 270 POLY-ALA.
 CC FT DOMAIN 273 284 POLY-GLY.
 CC FT CONFLICT 198 198 L -> I (IN REF. 2).
 CC FT CONFLICT 205 205 P -> PRRGAP (IN REF. 2).
 CC FT SEQUENCE 323 AA; 32406 MW; 598AE76CB512D716 CRC64;
 CC SQ
 CC Query Match 11.88; Score 68.5; DB 1; Length 323;
 CC Best Local Similarity 23.98; Pred. No. 12;
 CC Matches 33; Conservative 13; Mismatches 35; Indels 57; Gaps 7;
 CC
 CC QY 2 LGYSEPM-CANPLGLLGLRLPALSLPLV-----VTVGYMSAT----- 40
 CC Db 192 LAHSAPRLPTATAAAHAAHRAVHHPILPPAAAAAANAASASLPGSLSSV 251
 CC QY 41 -----PKKGLGECPPRRPPRYT-----GFTGDSGAKETYSQDKRSQGH-----TWCTLA 84
 CC Db 252 GSIRPPHGLKSPSAAAAAPLGGGGGGGGGGG-----HWGMPCPCSCQVP 301
 CC QY 85 LPHFWLTVGHLRNHVS 102
 CC Db 302 PPH-----HHVSA 309
 CC
 CC RESULT 12
 CC MTDI_ARATH STANDARD; PRT; 357 AA.
 CC ID MTDI_ARATH 095ZJ9;
 CC AC 002971; 095ZJ9;
 CC DT 01-NOV-1995 (Rel. 32, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Probable mannitol dehydrogenase 1 (EC 1.1.1.255) (NAD-dependent
 CC mannitol dehydrogenase 1).
 CC GN FL13-1 OR AT4G37980 OR F20D10.100.
 CC OS Arabidopsis thaliana (Mouse-ear cress).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 CC NCBI_Taxid=3702;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=cv. Columbia;
 CC RX MEDLINE=93099840; PubMed=1464303;
 CC Kiedrowski S., Kavalieck P., Hahlbrock K., Somschich I.E., Dangl J.L.;

RT "Rapid activation of a novel plant defense gene is strictly dependent
 RT on the Arabidopsis RPM1 disease resistance locus.";
 RT EMO J. 11:4677-4684(1992).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entlan K.-D., Terryn N.,
 RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weicheldartner M., de Simone V., Obermaier B., Maché R., Mueller T.,
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtner T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohnel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Gymonprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weltjens J., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weltzenger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirkes W.,
 RA Moolman P., Klein Lankhorst R., Rose M., Haut J., Koetter P.,
 RA Berner S., Hempel S., Feldpausch W., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gielen A., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McLay K., Mayes R.,
 RA Pettett A., Rajadream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fairman B., Grandtath K., Dauner D., Herzi A.,
 RA Neumann S., Argillon A., Vitale D., Liguori R., Piravandi E.,
 RA Maassen O., Outgley F., Clabaud G., Muendlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Cheifor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Perez-Perez A., Punelle B., Bent E., Johnson S., Tacon D., Jasse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Frances P., Bietke C.,
 RA Frishman D., Haese D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Shee P., Cordes M., Abu-Threideh J.,
 RA Stonehill T., Kallikl J., Graves T., Harmon G., Edwards J.,
 RA Latrellle P., Courtney L., Clout J., Abbott A., Scott K., Johnson D.,
 RA Mux P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drome K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Saby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Grant S., Shobhy N., Hasegawa A., Hameed A., Lochl M., Johnson A.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:769-777(1999).
 CC -1- FUNCTION: OXIDIZES MANNITOL TO MANNOSE. PROVIDES THE INITIAL STEP
 CC BY WHICH TRANSLOCATED MANNITOL IS COMMITTED TO CENTRAL METABOLISM
 CC AND, BY REGULATING MANNITOL POOL SIZE, IS IMPORTANT IN REGULATING
 CC SALT TOLERANCE AT THE CELLULAR LEVEL (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.
 CC -1- COFACTOR: ZINC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X67816; CAA48027.1; -
 CC EMBL: AL035538; CAB37538.1; -
 CC EMBL: AL161592; CAB80463.1; -

DR PIR: S28044; S28044.
 DR InterPro: IPR002328; ADH_zinc.
 DR InterPro: IPR002085; Adh_zn_family.
 DR Pfam: PF00107; adh_zinc.1.
 DR PROSITE: PS00059; ADH_ZINC; 1.
 KW Oxidoreductase; NAD; zinc; Multigene family.
 FT METAL 46 46 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 68 68 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 99 99 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 102 102 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 105 105 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 113 113 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 162 162 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 6 6 E -> Q (IN REF. 1).
 FT CONFLICT 17 17 E -> N (IN REF. 1).
 FT CONFLICT 20 20 I -> V (IN REF. 1).
 SQ SEQUENCE 357 AA; 38245 MM; 2C27B3C2BF030166 CRC64;
 Query Match 11.5%; Score 67; DB 1; Length 357;
 Best Local Similarity 25.2%; Pred. No. 19;
 Matches 35; Conservative 13; Mismatches 43; Indels 48; Gaps 6;
 QY 3 GISEPMPCAHPLGLFLGLHPALST---PLVVTAVGNSAPPKHGLECCPPAPPAVTG 58
 DB 133 GYSDHWVCAED---FLIRIPDNLPLDGAAPLLCAGVTVSPMKYHGLDK--PGMHGVVG 187
 QY 59 -----FTGDSGAKETV---SODKRSQGHWTWC----- 81
 DB 188 LGLGLHVAVKFAKAMKTQVTVSTSRKRDVAVRLGADAFVLRDPRKMDAMGTMDGI 247
 QY 82 --TLALPHPLTWVGHILRN 98
 DB 248 IDTVSATHTPLPLGLGLKN 266
 RESULT 13
 PGCA.CHICK STANDARD; PRT: 2109 AA.
 AC P07898; Q90991; Q90820; Q91047; Q90810;
 DT 01-ANG-1988 (Rel. 08, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPC).
 GN AGC1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-White Leghorn; TISSUE-Embryo;
 RX MEDLINE=94043149; PubMed=8226878;
 RA Li H., Schwartz N.B., Vertel B.M.;
 RT "CDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein and identification of a stop codon in the aggrecan gene associated with the chondrodystrophy, nanomelia.";
 RL J. Biol. Chem. 268:23504-23511(1993).
 RN [2]
 RP SEQUENCE OF 1042-1559 FROM N.A.
 RC TISSUE-Embryo;
 RX MEDLINE=90307744; PubMed=1694853;
 RA Krueger R.C., Jr., Fields T.A., Mensch J.R., Jr., Schwartz N.B.;
 RT "Chick cartilage chondroitin sulfate proteoglycan core protein. II. Nucleotide sequence of cDNA clone and localization of the S103L epitope.";
 RL J. Biol. Chem. 265:12088-12097(1990).
 RN [3]
 RP SEQUENCE OF 1-1855 AND 1893-2109 FROM N.A.
 RC TISSUE-Cartilage;
 RX MEDLINE=93111968; PubMed=1339285;
 RA Chandrasekaran L., Tanzer M.L.;

RT "Molecular cloning of chicken aggrecan. Structural analyses.";
 RL Biochem. J. 288:903-910(1992).
 RN [4]
 RP ERRATUM.
 RX MEDLINE=94107258; PubMed=8280087;
 RA Chandrasekaran L., Tanzer M.L.;
 RL Biochem. J. 296:885-887(1993).
 RN [5]
 RP SEQUENCE OF 1492-1610 FROM N.A.
 RC STRAIN-White Leghorn; TISSUE-Chondrocytes;
 RX MEDLINE=95128519; PubMed=7827752;
 RA Primorac D., Stover M.L., Clark S.H., Rowe D.W.;
 RT "Molecular basis of nanomelia, a heritable chondrodystrophy of chicken.";
 RL Matrix Biol. 14:297-305(1994).
 RN [6]
 RP SEQUENCE OF 1894-2109 FROM N.A.
 RX MEDLINE=89008500; PubMed=3170613;
 RA Tanaka T., Har-El R., Tanzer M.L.;
 RT "Partial structure of the gene for chicken cartilage proteoglycan core protein.";
 RL J. Biol. Chem. 263:15831-15835(1988).
 RN [7]
 RP SEQUENCE OF 1693-1855 AND 1893-2109 FROM N.A.
 RX MEDLINE=86259736; PubMed=3460082;
 RA Sai S., Tanaka T., Koshier R.A., Tanzer M.L.;
 RT "Cloning and sequence analysis of a partial cDNA for chicken cartilage proteoglycan core protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5081-5085(1986).
 CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGEOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
 CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
 CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS. N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -1- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF NANOMELIA, A LETHAL CONNECTIVE TISSUE DISORDER AFFECTING CARTILAGE DEVELOPMENT (CHONDRODYSSTROPHY) CHARACTERIZED BY SHORTENED AND MALFORMED LIMBS. AGGREGAN IS TRUNCATED AT ITS C-TERMINAL IN THE CS-2 BINDING DOMAIN AND IS NOT ANYMORE SECRETED FROM THE CHONDROCYTES.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 G-SH1 (SCR) DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib-sib.ch/announce/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L21913; AAB19128.1; -
 CC EMBL: M38187; AAA48731.1; -
 CC EMBL: M88101; -; NOT ANNOTATED_CDS.
 CC EMBL: S74657; AAC60751.1; -
 CC EMBL: S74656; AAC60751.1; JOINED.
 CC EMBL: J04028; AAA48719.1; -
 CC EMBL: M33993; AAA48720.1; -
 CC PIR: A25442; A25442.

Query	34	ACGYMSATPRHGLGECPP	-----APPAAVTGFGDGSAAKETYSQ-----	D 72
Db	650	AGGLAHAMVQGGIDRCYPCWLADGSLRYPYVSPRAC--	-GGDAPGVRTIYQHNNOTGFPD	706
QY	73	KRSQCHTQCTLALP	86	
Db	707	PLSRHHAFCRALP	720	
Query Match		11.5%	Score 67; DB 1; Length 2109;	
Best Local Similarity		29.7%	Pred. No. 1.2e+02;	
Matches	22;	Conservative	8; Mismatches	20; Indels 24; Gaps 3
FT	CARBOHYD	644	644	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	700	700	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	765	765	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	801	801	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VANOSPLIC	1856	1892	MISSING (IN ISOFORM 2).
FT	CONFLICT	362	362	E -> D (IN REF. 3).
FT	CONFLICT	601	601	G -> D (IN REF. 3).
FT	CONFLICT	1000	1000	P -> R (IN REF. 3).
FT	CONFLICT	1029	1029	A -> P (IN REF. 3).
FT	CONFLICT	1042	1043	VM -> PA (IN REF. 2).
FT	CONFLICT	1251	1251	E -> D (IN REF. 2 AND 3).
FT	CONFLICT	1387	1387	I -> T (IN REF. 5).
FT	CONFLICT	1590	1590	I -> V (IN REF. 5).
FT	CONFLICT	1594	1594	T -> S (IN REF. 5).
FT	CONFLICT	1602	1610	ITETSVREI -> VICRSYVL (IN REF. 5).
FT	CONFLICT	1603	1603	E -> A (IN REF. 3).
FT	CONFLICT	1672	1672	S -> G (IN REF. 3).
FT	CONFLICT	1796	1796	E -> G (IN REF. 3 AND 7).
RESULT 14				
CC12_SCHPO				
ID	CC12_SCHPO	STANDARD;	PRT; 1841 AA.	
AC	010059;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Cell division control protein 12.			
OS	CDC12 OR SPAC1F5.04C.			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomyces.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE-97258874; PubMed-9105045;			
RT	Chang F., Drubin D., Nurse P.;			
RT	"cdc12p, a protein required for cytokinesis in fission yeast, is a			
RT	component of the cell division ring and interacts with profilin."			
RL	J. Cell Biol. 137:169-182(1997).			
RN	[2]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=972;			
RC	MEDLINE-21648401; PubMed-11859360;			
RX	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,			
RX	Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,			
RA	Collins K., Brown D., Brown S., Chillingworth T., Churcher C., Fraser A.,			
RA	Brooks M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,			
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,			
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,			
RA	James K., Jones I., Jones M., Leather S., McDonald S., McLean J.,			
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,			
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,			
RA	Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,			
RA	Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,			
RA	Taylor K., Taylor R.G., Tiley A., Welsh S.V., Warren T., Whitehead S.,			
RA	Woodward J., Volkhardt G., Aert R., Robben J., Gromoprez B.,			
RA	Weijens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,			
RA	Gabel K., Fuchs W., Fritz C., Holzer E., Moesti D., Hilbert H.,			
RA	Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,			

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Gaubert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochez M., Gallardin C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,
RA "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- FUNCTION: PLAYS A ROLE IN THE CELL CYCLE. INVOLVED IN CYTOKINESIS.
CC COMPONENT OF THE CELL DIVISION RING; INTERACTS WITH PROFILIN.
CC -1- SIMILARITY: TO YEAST BNI1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z68136; CA92232.1; -
DR InterPro: IPR003104; FH2.
DR Pfam: PF02181; FH2; 1.
DR SMART: SM00498; FH2; 1.
KW Cell division; cell cycle.
SQ SEQUENCE 1841 AA; 207598 MW; E1AA42731FCD37C5 CRC64;
OY Query Match 11.4%; Score 66.5; DB 1; Length 1841;
DB Best Local Similarity 33.3%; Pred. No. 1.1e+02;
Matches 23; Conservative 6; Mismatches 27; Indels 13; Gaps 4;
OY 7 PMPCAHPLGLFLHPLSLPLV---VTAGVMSAPFKHGLEQCPAPAPVNGFTGDS 63
DB 912 PLPKTSINTNT---SHPD-SVNIYANDTSVAGVMPAPFP-----PPPPPLVSAAGCK 961
OY 64 GAKETVSOD 72
DB 962 FVSPAVSNM 970
RESULT 15
ID GP62_HUMAN STANDARD; PRT; 368 AA.
AC O982J7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable G protein-coupled receptor GPR62.
GN GPR62.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21105913; PubMed=11165367;
RA Lee D.K., George S.R., Cheng R., Nguyen T., Liu Y., Brown M.,
RA Lynch K.R., O'Dowd B.F.;
RT "Identification of four novel human G protein-coupled receptors
RT expressed in the brain."
RL Brain Res. Mol. Brain Res. 86:13-22(2001).
CC -1- FUNCTION: Orphan receptor.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in brain; detected in the basal
CC forebrain, frontal cortex, caudate, putamen, thalamus and
CC hippocampus.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF317653; AAK12638.1; -
DR Genew; HGNC:13301; GPR62.
DR MIM: 606917; -
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm_1; 1
DR PRINTS: PR00237; GPCR_Rhodopsn.
DR PROSITE: PS00237; G_PROTEIN_RECIP_F1_1; FALSE NEG.
DR PROSITE: PS50262; G_PROTEIN_RECIP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 18
FT TRANSMEM 19 39
FT DOMAIN 40 53
FT TRANSMEM 54 74
FT DOMAIN 75 91
FT TRANSMEM 92 112
FT DOMAIN 113 129
FT TRANSMEM 130 150
FT DOMAIN 151 177
FT TRANSMEM 178 198
FT DOMAIN 199 239
FT TRANSMEM 240 260
FT DOMAIN 261 272
FT TRANSMEM 273 293
FT DOMAIN 294 368
FT TRANSMEM 369 385
FT CARBOHYD 3 8
FT CARBOHYD 3 8
SQ SEQUENCE 368 AA; 37628 MW; 9CFF95298D12C75 CRC64;
OY Query Match 11.3%; Score 66; DB 1; Length 368;
DB Best Local Similarity 26.9%; Pred. No. 24;
Matches 32; Conservative 6; Mismatches 37; Indels 44; Gaps 5;
OY 7 PMPC-----AHPGLGLFLGL-----HPALSLPLVTVVAGVMS 38
DB 85 PAPCRAPFLSNALLPACTLGVALLGLARYLRYHPLRPGSRPPVLTAVVMAAGLIG 144
OY 39 ATPKHGLQCPAPPPA---VTGFTGDSGAKETVSQDKRSQGHWTCTLALPMPWLTWG 94
DB 145 ALSLIG---PPAPPPAPARCSYLAGLGGPPAPL-----WALLAPLPLALLIG 190
Search completed: June 17, 2003, 10:35:27
Job time : 24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:32:51 : Search time 35 Seconds

(Without alignments)
258.823 Million cell updates/sec

Title: US-09-955-807-2

Sequence: 1 MGVSEPMPCAHPLGLGLG.....PHFWLWVGHILRNHVSASH 105

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	13.7	264	2 A49149	mesoderm developme
2	73.5	12.6	334	2 G02409	protein kinase C-b
3	72.5	12.5	554	2 F86244	hypothetical prote
4	72	12.4	575	2 T11753	mullerian inhibiti
5	72	12.4	1089	2 T36663	protein kinase, tr
6	71	12.2	110	2 JN0707	proteinaceous alph
7	71	12.2	1964	2 T09059	notch4 - mouse
8	70	12.0	1616	2 G70668	polyketide synthas
9	69.5	11.9	434	2 A75450	conserved hypothet
10	69.5	11.9	860	2 C82750	mannosyltransferas
11	69	11.9	287	2 T22637	hypothetical prote
12	69	11.9	328	2 C87451	hypothetical prote
13	68.5	11.8	202	2 S76283	hypothetical prote
14	68.5	11.8	399	2 AD2559	hypothetical prote
15	68	11.7	1400	2 B70963	hypothetical prote
16	67.5	11.6	1189	2 JC6118	SH2-containing ino
17	67	11.5	357	2 T05625	climamyl-alcohol d
18	67	11.5	980	2 S54986	regulatory protein
19	67	11.5	2109	1 T50421	aggrean precursor
20	66.5	11.4	906	2 A71438	probable resistanc
21	66.5	11.4	1841	2 T38091	cell division cont
22	66	11.3	262	2 D97681	hypothetical prote
23	66	11.3	686	1 VGBENA	conserved hypothet
24	66	11.3	262	1 AC2906	glycoprotein H pre
25	66	11.3	899	2 G02428	subtilisin-like pr
26	66	11.3	915	1 A48225	subtilisin-like pr
27	66	11.3	915	2 B48225	subtilisin-like pr
28	66	11.3	915	2 JC6148	subtilisin-like pr
29	66	11.3	920	2 JC7313	aryl hydrocarbon r

30	66	11.3	1548	2 S34583	serine proteinase
31	66	11.3	3869	2 A48205	All-1 protein +GTR
32	65.5	11.3	296	2 T24827	hypothetical prote
33	65.5	11.3	346	2 T21552	hypothetical prote
34	65.5	11.3	348	2 T02462	probable AT-hook D
35	65.5	11.3	422	2 T09120	basic leucine zip
36	65.5	11.3	439	2 D83345	L-sorbose dehydr
37	65.5	11.3	478	2 A49228	trypsin-like prote
38	65.5	11.3	556	2 B87193	hypothetical prote
39	65.5	11.3	2297	2 T34918	polyketide synthas
40	65.5	11.3	2831	2 T31419	cyclic beta 1-2 gl
41	65.5	11.3	2867	2 AC3481	cellulose-phospho
42	65.5	11.3	3685	1 A27605	dystrophin, muscle
43	65	11.2	576	1 B70558	probable ABC trans
44	65	11.2	607	2 T32968	hypothetical prote
45	65	11.2	686	1 VGBEPK	glycoprotein H pre

ALIGNMENTS

```

RESULT 1
A49149
mesoderm development regulatory protein Sna - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 01-Dec-2000
C:Accession: A49149; S24674; A49124
R:Nieto, M.A.; Bennett, M.F.; Sargent, M.G.; Wilkinson, D.G.
Development 116, 227-237, 1992
A:Title: Cloning and developmental expression of Sna, a murine homologue of the Dros
A:Reference number: A49149; MUID:93130772; PMID:1483390
A:Accession: A49149
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-264 <SAR>
A:Cross-references: GB:X67253; NID:954120; PIDN:CAA47675.1; PID:954121
A:Experimental source: embryo
A>Note: sequence extracted from NCBI backbone (NCBIN:122886, NCBIPI:122887)
R:Sargent, M.G.
submitted to the EMBL Data Library, July 1992
A:Reference number: S24674
A:Accession: S24674
A:Molecule type: mRNA
A:Residues: 1-264 <SAR>
A:Cross-references: EMBL:X67253; NID:954120; PIDN:CAA47675.1; PID:954121
R:Smith, D.E.; Franco del Amo, F.; Gridley, T.
Development 116, 1033-1039, 1992
A:Title: Isolation of Sna, a mouse gene homologous to the Drosophila genes snail and
A:Reference number: A49124; MUID:93201990; PMID:1295727
A:Accession: A49124
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-218, 'D', 220-264 <SMI>
A:Cross-references: GB:M95604; NID:9429187; PIDN:AAA03481.1; PID:9429188
A:Experimental source: C57BL/6, embryo
A>Note: sequence extracted from NCBI backbone (NCBIN:127997, NCBIPI:127998)

Query Match      13.7% Score 80; DB 2; Length 264;
Best Local Similarity 24.5% Pred. No. 1.4;
Matches 46; Conservative 13; Mismatches 37; Indels 92; Gaps 11;

OY 4 YSEPMPCAHPLGLGLG-----LHRLSLP-----LVVTYAGVMSAT-----PK- 42
DB 31 FOQPYDQAH-----LLAIPPEVLPNPAASLPILWDSLIVQVPAVAMATLPRESPKA 85
OY 43 -----HGLQCPAPP-PAVTGPTGDSG-----AKETVS 70
DB 86 VELTSLSDSDSGKSSQPPSPSPSPSSSTSSASLEAFIAPGIGOLPKQLAKRSVA 145
OY 71 QDKRS-----OGH-----TWCTLALPHFWLTVGHLRNHVS 101
DB 146 KDQSRKIFNCKYCNKEYLSIGALKMHIRSHTLPCVCTGCKAKSRPMLL-OGHVRHTG 204

```



```
C:Species: Streptomyces nitrosporeus  
C>Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_Change 28-May-1999  
C:Accession: JN0707  
R:Sunltani, J.; Kawaguchi, T.; Hattori, N.; Murao, S.; Arai, M.  
Biosci. Biotechnol. Biochem. 57, 1243-1248, 1993  
A>Title: Molecular cloning and expression of proteinaceous alpha-amylase inhibitor gene  
A:Reference number: JN0707; MUID:93379351; PMID:7764011  
A:Accession: JN0707  
A:Molecule type: DNA  
A:Residues: 1-110 <SD>  
A:Cross-references: GB:S65457; NID:g451243; PIDN:AAC60452.1; PID:g451244  
C:Comment: This enzyme is active specifically against animal amylases from animals but has no activity towards plant or fungal amylases.  
C:Superfamily: Streptomycetes alpha-amylase inhibitor  
C:Keywords: alpha-amylase inhibitor  
F:1-33/Domain: signal sequence  
F:34-110/Product: proteolnaceous alpha-amylase Inhibitor #status predicted <SIG>  
  
Query Match          12.2%   Score 71; DB 2; Length 110;  
Best Local Similarity 29.9%; Pred. No. 4.3;  
Matches      26; Conservative       7; Mismatches    44; Indels     10; Gaps    4;
```

OY		21 LHPALSLPLVTVAGVMSAPFKHGLEDCPPAPPA-VTGFTj-----DSGAKETVSQDK 73
Dg		8 ISPALTTLATVF-AAGLMTLLAAPSAAMAATCAPACAVESFQSWMRTDVTHNGCSSETVASVT 66
OY		74 RSQGHTW--CTLLAPHPMLTGWGHLRN 98
Dg		67 EYTNGQMARPCTVIOPDGMAITPAGYGIN 93

C:Date: 17-Jul1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: G70668; S73075
R:COLE, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
J.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajadaram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A70500; MUID: 98295987; PMID: 9634230
A:Accession: G70668
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1616 <COL>
A:Cross-references: GB:283858; GB:AL123456; NID:g3261675; PIDN:CAAB06103.1; PID:g17811
A:Experimental source: Strain H37Rv
R:Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, September 1994
A:Description: *Mycobacterium tuberculosis* cosmid tbc2.
A:Reference number: S73053
A:Accession: S73075
A:Molecule type: DNA
A:Residues: 'MNGAMPVARKSTEL', 48-917, 'RSEARSRCGSGSRAPAPPGPPIG', 'PMPPIIPEVPPAP
A:Cross-references: EMBL:U00024; NID:g560506; PIDN:AAA50928.1; PID:g560507
C:Genetics:
A:Gene: pks1
C:Superfamily: *Mycobacterium tuberculosis* probable polyketide synthase pks1; acyl car
riery
C:Keywords: carrier protein
C:KeyWords/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
;84-362/Domain: [acyl-carrier-protein]

RESULT 7
T09059
notch4 - mouse
C.Species: Mus musculus (house mouse)
C.Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jan-2000
C.Accession: T09059
R.Röwen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.; Sch
submitted to the EMBL Data Library, October 1997
A.Description: Sequence of the mouse major histocompatibility locus class III region.
A.Reference number: Z16543
A.Accession: T09059
A.Status: preliminary; translated from GR/EMBL/DBD
A.Molecule type: DNA
A.Residues: 1-1964 <ROW>
A.Cross-references: EMBL:AF030001; NID:q2564945; PID:g1564947
C.Genetics:
A:Gene: notch4
A:Map position: 17
A:Introns: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 67
1679/3; 1729/1; 1761/3
C.Superfamily: unassigned EGF-related proteins; EGF homology
C.Keywords: receptor; signal transduction
;:514-545/Domain: EGF homology <EGF>

```

Query Match          12.0% Score 70; DB 2; Length 1616;
Best Local Similarity 34.6%; Pred. No. 88;
Matches 28; Conservative 10; Mismatches 35; Indels 8; Gaps 4;

QY      19 LGHPALSLPIVTVAGVMSATPKHGECPAPPAPVVTGFTGDSGARETVSOD---KRS 75
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db      950 LGMTPGQAPPLGAGACAGVLETT---GPETVDLAVGDAMGMLGAGGLAVDDOOLYTRVP 1006
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

QY      76 QGHWC-T-LALPHPLT-WWG 94
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Db      1007 QGMSFQAQAAPVVFLTWYG 1027

RESULT 9
A75450
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: A75450
R:White, O.; Eisten, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
```

```
Oy      3 GYSEPMCAHILGLFLGLHPALSLPLVWVAGWSMTAPKRGLE-----CCPPAPP----- 53
        |   |||
Db      1129 GCGRPSRLH-----NCTCTETPGELGNCFQCTCPDPSPGRCQ 1167
               :
Oy      54 -PAVTGFTGDSGAKEIVSQQRRKSQGHTW----CTLALPHEW 89
        |   |   |   |   |   |   |   |   |   |   |   |
Db      1168 RFGASGGEG-RGGDGTCDACGCCGGGDMGDGDCSLGVPDFW 1207
```

A:Title: Genome sequence of the radiorresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: A75450
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1434 <RND>
A:Cross-references: GB:AE001952; GB:AE000513; NID:96458725; PIDN:AAF10584.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1009
A:Map position: 1

RESULT 8
G70668
polyketide synthase pksI - Mycobacterium tuberculosis (strain H37RV)
N:Alternate names: polyketide synthase pks002c
C:Species: Mycobacterium tuberculosis

```

07      26 SLPLVTVYACVMSATPKHGLEQC--PPAPPAVTGTFGDSGAKEVYSQDK----- 73
          ::::| | ||||| :|| : : ||:
Best Local Similarity 29.3%; Pred. No. 25;
Matches 27; Conservative 9; Mismatches 37; Indels 19; gaps 5;

```

```

Db      250 NVPATVGVAPAPQPTBPAPTBPAPBPATPGSGYTWYPESNNRNASADELEIIRLTNEA   309
Oy      74 RSOCHTWC--TLALPHPLTWVGHLL----RH  99
        ||| | | | | | | | | | | | | | | |
Db      310 RARGAT-CGTTTPYAPAPALRNDDLAHASNHH 340

RESULT 10
C82750
manosyltransferase XF0867 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: C82750
R:Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences
Nature 406, 151-157, 2000
A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A>Note: for a complete list of authors see reference number A59328 below
A:Accession: C82750
A>Status: preliminary
A:Molecule type: DNA
A:Locus: preliminary
A:Cross-references: GB:A0003928; GB:AE003849; MID:g9105798; PIDN:AA83697.1; GSPDB:GN0001
R:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; B
Briotes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, I
as-Neto, E.; Docena, C.; El-Dorcy, H.; Facinanci, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.;
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasab
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vetlore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0867

Query Match          11.9%, Score 69.5; Db 2; Length 860;
Best Local Similarity 27.8%; Pred. No. 51;
Matches 35; Conservative 11; Mismatches 47; Indels 33; Gaps 7;

Oy      5 SEPP--CAHPGLGFLGLHPALSLPLYVTYAGWSATPKHGEECPAPPRAVATGTGD 62
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      502 STPPPGVCIVPVYLREGGYHYARS--FTAOVLGIASENPPEGIVMYVPGDV-----EVGL 554

Oy      63 SGARETVS-----ODKRSQS-----HTWCTALP---HP-----MLTWVGHLRN 98
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      555 DMAEPAINMAARALQDMRRRGVATCFYVHDLLPMTLPDCHFPYSRYNLFEOMLRTVAHLAD 614
        ||| | | | | | | | | | | | | | | | | | | | | | | | | |

Oy      99 HVSSAS 104
        : : |
Db      615 AIACIS 620

RESULT 11
T22637
hypothetical protein F54C9.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T22637
R:Sims, M.
submitted to the EMBL Data Library, June 1995
A:Reference number: Z19591
A:Accession: T22637
A>Status: preliminary; translated from GR/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1..287 <WIL>
A:Cross-references: EMBL:T249967; PIDN:CAA90250.1; GSPDB:GN00020; CESP:F54C9.4
A:Experimental source: clone F54C9

```

```
C|Genetics: A:Gene: CESP:F54C9.4  
A:Map position: 2  
A:Introns: 46/73  
C|Superfamily: unassigned collagens
```

```
Query Match          11.9%; Score 69; DB 2; Length 287;  
Best Local Similarity 40.5%; Pred. No. 18;  
Matches      17; Conservative    4; Mismatches   15; Indels       6; Gaps        2;
```

```
OY           48 CPPAP--PPAVTGFTGDSGAKEIVSODKRSQGHMTCTALAPH 87  
            |||| | | | | | | | | | | : | | |  
DB           97 CPFGPBPGPGLPGGRGDGP-----SGDKRGDAGASGVSLAYTH 134
```

```
RESULT 12  
C87451
```

```
hypothetical protein CC1629 [imported] - Caulobacter crescentus  
C|Species: Caulobacter crescentus  
C|Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001  
C|Accession: C87451  
R:Nierman, W.C.; Feldbljum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
B.J.; Labib, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A>Title: Complete Genome Sequence of Caulobacter crescentus.  
A|Reference number: A87249; MUID:21173698; PMID:11259647  
A|Accession: C87451  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-328 <STO>  
A:Cross-references: GB:AEO05673; NID:g13423031; PIDN:AAK23607.1; GSFPDB:GN00148  
C|Genetics:  
A:Gene: CC1629
```

```
Query Match          11.9%; Score 69; DB 2; Length 328;  
Best Local Similarity 28.9%; Pred. NO. 21;  
Matches     26; Conservative    8; Mismatches   22; Indels     34; Gaps      5;
```

```
OY           13 PLGLELLGLHPALSLPL-----VTVVGVMASATP-----KHGLEOCPPA 51  
            |:|:| | | | | | | | | | | | | | | | | | | | | | | |  
DB           55 PVGIQLXSLGDLAKELDAQLATAVKIGFRTVELAGLGRTPALRALPDFRHGL-VCPSPA 113
```

```
OY           52 --PPPAVTGFEGD-----SGAEKV 69  
            + | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB           114 HISPKGPNGEFGDLVLKLDELHYLGAKSAI 143
```

```
RESULT 13  
S76283
```

```
hypothetical protein sll0756 - Synechocystis sp. (strain PCC 6803)  
C|Species: Synechocystis sp.  
A|Variety: PCC 6803  
C|Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000  
C|Accession: S76283  
R:Kaneko, T.; Satoh, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O.K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas  
DNA Res. 3, 109-116, 1996  
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechoys  
s.
```

```
A|Reference number: S74322; MUID:97061201; PMID:8905231  
A|Accession: S76283  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-202 <KAN>  
A:Cross-references: EMBL:D64000; GB:AB003339; NID:g1001484; PIDN:BAA10135.1; PID:g100  
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C|Genetics:  
A|Start codon: GTG  
C|Superfamily: Synechocystis hypothetical protein sll0756
```

```
Query Match          11.8%; Score 68.5; DB 2; Length 202;  
Best Local Similarity 31.9%; Pred. No. 14;
```

Matches 29; Conservative 7; Mismatches 20; Indels 35; Gaps 6;
 QY 1 MGVSEP-----MCAHPLIGLFIILGLHPALSLPLVTVAGVMSATPKHGLEQCPAPPAPPA 55
 Db 29 MINPFWLMSMPLTPPLTLAMAGT--ALSLG-----ATPGHG-ETCVPIP--- 73
 QY 56 VTGTGDSGAKETVSQDKRSQGHWTCTALP 86
 Db 74 LVGGGNSVTK-----TVAVP 89

RESULT 14

AD2559
 hypothetical protein al18067 [imported] - Nostoc sp. (strain PCC 7120) plasmid pcc7120ge
 C:Species: Nostoc sp.
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AD2559
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD2559
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-399 <KUR>
 A:Cross-references: GB:AP003603; PIDN:BA87397.1; PID:917134840; GSPDB:GN00182
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: al18067
 A:Genome: plasmid

Query Match 11.8%; Score 68.5; DB 2; Length 399;
 Best Local Similarity 26.9%; Pred. No. 29;
 Matches 18; Conservative 15; Mismatches 13; Indels 21; Gaps 3;

QY 23 PALSLPLVTVAGVMSATPKHGLEQCPAPPAPVATGFTGDSGAKETVSQDKRSQGHWTCT 82
 Db 63 PSISIP-----TASATPKNSL-----PTSVNASKNNKASSTITQERK-----T 101

QY 83 LALPHPW 89
 Db 102 VLPPNPW 108

RESULT 15

B70963
 hypothetical protein RV0236c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 28-Jul-2000
 C:Accession: B70963
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skilton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: B70963
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1400 <COL>
 A:Cross-references: GB:Z92669; GB:AL123456; NID:93242271; PIDN:CAB07017.1; PID:93242273
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV0236c
 C:Superfamily: Mycobacterium leprae probable integral membrane protein

Query Match 11.7%; Score 68; DB 2; Length 1400;
 Best Local Similarity 26.8%; Pred. No. 1.2e+02;
 Matches 33; Conservative 9; Mismatches 37; Indels 44; Gaps 7;

QY 9 PCAHPLIGF-----ILGLH---PALSLPLVTVAGVMSATPKHGLEQCPAPPAPPA-VT 57
 Db 336 PVAPHPVQAFLLDAAGTPIRLNRYKVGVPVIRLPLVLGLAQLSTRVPLRG-----SAPRPAMLR 390
 QY 58 GFTGDSGAKETVSQDKR-----SQGHWTCTL-----ALPHPWLTWVGH 96
 Db 391 AFAPHP-----ERDKRVAVAVALTALMSTSLAMTGRVAPPGTFGALPYQWQEAADWL 443
 QY 97 RNH 99
 Db 444 RTH 446

Search completed: June 17, 2003, 10:37:41
 Job time : 42 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:33:46 ; Search time 25 Seconds
(without alignments)
118.823 Million cell updates/sec

Title: US-09-955-807-2
Perfect score: 582
Sequence: 1 MLCSEPMPCAHPLGLFLILG.....PMPWLTVYGHILRNHYSSASH 105

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PC10S.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71	12.2	1964	US-09-467-997-1	Sequence 1, Appl1
2	67.5	11.6	1185	US-08-664-962B-2	Sequence 2, Appl1
3	67.5	11.6	1185	US-09-311-743-2	Sequence 2, Appl1
4	66	11.3	306	US-08-368-853-13	Sequence 13, Appl1
5	66	11.3	306	US-08-525-940-13	Sequence 13, Appl1
6	66	11.3	306	US-08-976-838-13	Sequence 13, Appl1
7	66	11.3	799	US-08-525-940-23	Sequence 23, Appl1
8	66	11.3	799	US-08-976-838-23	Sequence 23, Appl1
9	66	11.3	881	US-08-525-940-21	Sequence 21, Appl1
10	66	11.3	881	US-08-976-838-21	Sequence 21, Appl1
11	66	11.3	915	US-08-525-940-18	Sequence 18, Appl1
12	66	11.3	915	US-08-976-838-18	Sequence 18, Appl1
13	66	11.3	915	US-09-214-555B-2	Sequence 2, Appl1
14	66	11.3	915	US-09-214-555B-7	Sequence 7, Appl1
15	64.5	11.1	326	US-08-053-867A-2	Sequence 2, Appl1
16	64.5	11.1	326	US-08-053-867A-8	Sequence 8, Appl1
17	64.5	11.1	2972	US-09-579-181-2	Sequence 2, Appl1
18	64.5	11.1	3118	US-09-579-181-1	Sequence 1, Appl1
19	64	11.0	380	US-09-026-587-4	Sequence 4, Appl1
20	64	11.0	380	US-09-227-420-4	Sequence 4, Appl1
21	63.5	10.9	369	US-07-688-352C-32	Sequence 32, Appl1
22	63.5	10.9	369	US-08-474-379C-32	Sequence 32, Appl1
23	63.5	10.9	369	US-09-146-249A-32	Sequence 32, Appl1
24	63.5	10.9	369	US-08-206-188B-32	Sequence 32, Appl1
25	63.5	10.9	369	PCR-US91-02714-31	Sequence 31, Appl1
26	63.5	10.9	380	US-09-673-397-2	Sequence 2, Appl1
27	63.5	10.9	380	US-09-673-397-6	Sequence 6, Appl1

28	63.5	10.9	380	US-09-673-397-8	Sequence 8, Appl1
29	63.5	10.9	1447	US-08-540-406-19	Sequence 19, Appl1
30	63.5	10.9	1447	US-08-656-055-19	Sequence 19, Appl1
31	63.5	10.9	1447	US-08-954-668-19	Sequence 19, Appl1
32	63.5	10.9	1447	US-09-268-140-5	Sequence 5, Appl1
33	63.5	10.9	1447	US-08-918-658-19	Sequence 19, Appl1
34	63.5	10.9	1447	PCR-US95-13233-19	Sequence 19, Appl1
35	63	10.8	1434	US-08-540-406-10	Sequence 10, Appl1
36	63	10.8	1434	US-08-656-055-10	Sequence 10, Appl1
37	63	10.8	1434	US-08-954-668-10	Sequence 10, Appl1
38	63	10.8	1434	US-08-918-658-10	Sequence 10, Appl1
39	63	10.8	1434	PCR-US95-13233-10	Sequence 10, Appl1
40	63	10.8	2930	US-09-417-822-2	Sequence 2, Appl1
41	62.5	10.7	113	US-08-241-853-15	Sequence 15, Appl1
42	62.5	10.7	113	US-08-850-917-15	Sequence 15, Appl1
43	62.5	10.7	114	US-08-844-056-4	Sequence 4, Appl1
44	62.5	10.7	418	PCR-US94-01321-72	Sequence 72, Appl1
45	62.5	10.7	443	US-09-161-994A-3	Sequence 3, Appl1

ALIGNMENTS

```
RESULT 1
US-09-467-997-1
; Sequence 1, Application US/09467997
; Patent No. 6379925
; GENERAL INFORMATION:
; APPLICANT: Kitejowski, Jan
; APPLICANT: Uytendaele, Hendrik
; TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION
; FILE REFERENCE: 53863-A-PCR-US
; CURRENT APPLICATION NUMBER: US/09/467, 997
; CURRENT FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1964
; TYPE: PRT
; ORGANISM: mouse
US-09-467-997-1
Query Match      12.2%; Score 71; DB 4; Length 1964;
Best Local Similarity 24.8%; Pred. No. 25;
Matches 25; Conservative 4; Mismatches 36; Indels 36; Gaps 5;

QY      3 GTSEPMPCAHPLGLFLILGHPALSLPLVYTVAGVMSATPKHGLE-----QCPAPP----- 53
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1129 GCGPPSPCLH-----NGCTETFGILGNPGFOCTCPDPSGPRQO 1167
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      54 -PAVTFGDSGAKETVSQDKRSQCHTW----CTLALPHPW 89
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1168 RFGASGCGE-RGGDGTCDAGCGSGGDMWDGDCSLGVPDPW 1207
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
US-08-664-962B-2
; Sequence 2, Application US/08664962B
; Patent No. 6218162
; GENERAL INFORMATION:
; APPLICANT: Krystal, Gerald
; TITLE OF INVENTION: SH2-CONTAINING INOSITOL-PHOSPHATASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MERCHANT & GOULD
; STREET: 3100 No. 6218162west Center, 90 South Seventh Street
; CITY: Minneapolis
; STATE: Minnesota
; COUNTRY: U.S.A.
; ZIP: 55402-4131
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/664,962B
FILING DATE: 14-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mueller, Douglas P.
REGISTRATION NUMBER: 30,300
REFERENCE/DOCKET NUMBER: M6G 7933.49-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1185 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-664-962B-2

Query Match 11.6%; Score 67.5; DB 4; Length 1185;
Best Local Similarity 28.9%; Pred. No. 34;
Matches 22; Conservative 10; Mismatches 29; Indels 15; Gaps 4;

QY 6 EPMCAHPLGLFLGL-HPALSLPLVYVAVMSATPKHGLECCP--APPAYTGFTG 61
DB 1041 EPPPCPP-----GISSPSIVLPKQEVESVKGTS-----KQAPVPLGPTPRIRSFCT 1089

QY 62 DSGAKETVSQDKRSOG 77
DB 1090 SSSAEGRMVSGDKSQG 1105

RESULT 3

US-09-311-743-2
Sequence 2, Application US/09311743
Patent No. 6238903

GENERAL INFORMATION:
APPLICANT: Krystal, Gerald
TITLE OF INVENTION: SH2-CONTAINING INOSITOL-PHOSPHATASE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERESKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/311,743
FILING DATE: 14-May-1999
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Michelle
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 7771-32
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-364-7311
TELEFAX: 416-361-1398

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1185 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-311-743-2

Query Match 11.6%; Score 67.5; DB 4; Length 1185;
Best Local Similarity 28.9%; Pred. No. 34;
Matches 22; Conservative 10; Mismatches 29; Indels 15; Gaps 4;

QY 6 EPMCAHPLGLFLGL-HPALSLPLVYVAVMSATPKHGLECCP--APPAYTGFTG 61
DB 1041 EPPPCPP-----GISSPSIVLPKQEVESVKGTS-----KQAPVPLGPTPRIRSFCT 1089

QY 62 DSGAKETVSQDKRSOG 77
DB 1090 SSSAEGRMVSGDKSQG 1105

RESULT 4

US-08-368-852-13
Sequence 13, Application US/08368852
Patent No. 5691183

GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/368,852
FILING DATE: 05-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: 2848-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223

INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 306 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-368-852-13

Query Match 11.3%; Score 66; DB 1; Length 306;
Best Local Similarity 27.5%; Pred. No. 9.8;
Matches 25; Conservative 16; Mismatches 26; Indels 24; Gaps 5;

QY 37 MSATPKRG-----LECPAPAPPPVGTGDSGAKETVSQDKR---SQGHTWCTALP-- 86
DB 200 ISSIASGKKPMYLEEC--SSLATVYSSGESYDKKITITDLOQCTDNTHTGSASAPMA 257

QY 87 -----HPWLTV--VGLRNHVSASH 105
DB 258 AGIIALALEANPFLTRMDVGVIVRISRAGH 288

RESULT 5

US-08-525-940-13
Sequence 13, Application US/08525940
Patent No. 586351

GENERAL INFORMATION:

```

APPLICANT: Franzusoff, Alex
APPLICANT: Miranda, Luis R.
APPLICANT: Wolf, Joseph R.
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
TITLE OF INVENTION: ENCODING SAID PROTEASES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,940
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/368,852
FILING DATE: 01-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/088,322
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 306 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-525-940-13

Query Match      11.3%; Score 66; DB 2; Length 306;
Best Local Similarity 27.5%; Pred. 9.8;
Matches   25; Conservative    16; Mismatches     26; Indels     24; Gaps     5;

CY          37 MSATPKRG-----TEOCPAPPVAVTGPTGSGAKETVSODKR--SSQHNMCITLALP-- 86
             :|::||::||::||::||::||::||::||::||::||::||::||::||:
DB          200 ISSIAEGSKRRWYLEEC--STLATIYYSGGSYDKKITTLRLRQRCTDNNHTGTASADMA 257
               ||:|||::||::||::||::||::||::||::||::||::||::||::||
              HPWLTV--VGHLRNHYVSASH 105
              ::|||::||::||::||::||::||::||::||::||::||::||::||
QB          258 AGITALLEANPFILTRDVQHIVRTSPAGH 288

RESULT 6
US-08-976-838-13
Sequence 13, Application US/08976838
Patent No. 5981259
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,838
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C2
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 306 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-976-838-13

Query Match 11.3%, Score 66; DB 2; Length 306;
Best Local Similarity 27.5%; Pred. No. 9.8;
Matches 25; Conservative 16; Mismatches 26; Indels 24; Gaps 5

OY 37 MSATPKHG-----EEOCPAPPPAVTGTGDSGAKEIVSODK---SQGHFWCTLALP-- 86
 |::| ::|| ::| : || ::| : || ::| : || ::| : ||
Db 200 ISSIAESGRKKRWYLEEC--STLATYTSSGESDYDKKITITDLRQCRDINHIGTSASAEPA 257

OY 87 -----HPWLTV--VGLLRNHYVSASH 105
 ::||| ::| : |||
Db 258 AGITALLEANPFLLWRDVQHVIYRTSRAGH 288

RESULT 7
US-08-525-940-23
Sequence 23, Application US/08525940
Patent No. 5866351
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
APPLICANT: Miranda, Luis R.
APPLICANT: Wolf, Joseph R.
TITLE OF INVENTION: CD4+ T-LYMPCYOYE PROTEASES AND GENES
TITLE OF INVENTION: ENCODING SAID PROTEASES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.

ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,940
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/368,852
FILING DATE: 01-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/088,322
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C1

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-940-23

Query Match 11.3%; Score 66; DB 2; Length 799;
Best Local Similarity 27.5%; Pred. No. 31;
Matches 25; Conservative 16; Mismatches 26; Indels 24; Gaps 5;

OY 37 MSATPKHG-----LEQCPAPPPAVTGTGDSGAKETVSODKR---SQGHTWCTLALP-- 86
Db 221 ISSTAESGKKPWYLEEC--SSTLATYSSGESYDKKITITDLQRCRDNDHTGTSASAPMA 278
OY 87 -----HPWLTV--VGLLRNHVSSASH 105
Db 279 AGIITALLEANPFLTRDVOHVIVRTSRAGH 309

RESULT 8
US-08-976-838-23
; Sequence 23, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
; NUMBER OF INVENTION: MOLECULES
; CORRESPONDENCE ADDRESSES: 31
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,838
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-976-838-23

Query Match 11.3%; Score 66; DB 2; Length 799;
Best Local Similarity 27.5%; Pred. No. 31;
Matches 25; Conservative 16; Mismatches 26; Indels 24; Gaps 5;

OY 37 MSATPKHG-----LEQCPAPPPAVTGTGDSGAKETVSODKR---SQGHTWCTLALP-- 86
Db 221 ISSTAESGKKPWYLEEC--SSTLATYSSGESYDKKITITDLQRCRDNDHTGTSASAPMA 278
OY 87 -----HPWLTV--VGLLRNHVSSASH 105
```

```

Db 279 AGIITALLEANPFLTRDVOHVIVRTSRAGH 309

RESULT 9
US-08-525-940-21
; Sequence 21, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
; NUMBER OF INVENTION: ENCODING SAID PROTEASES
; CORRESPONDENCE ADDRESSES: 25
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,852
; FILING DATE: 01-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/088,322
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 881 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-940-21

Query Match 11.3%; Score 66; DB 2; Length 881;
Best Local Similarity 27.5%; Pred. No. 35;
Matches 25; Conservative 16; Mismatches 26; Indels 24; Gaps 5;

OY 37 MSATPKHG-----LEQCPAPPPAVTGTGDSGAKETVSODKR---SQGHTWCTLALP-- 86
Db 303 ISSTAESGKKPWYLEEC--SSTLATYSSGESYDKKITITDLQRCRDNDHTGTSASAPMA 360
OY 87 -----HPWLTV--VGLLRNHVSSASH 105
Db 361 AGIITALLEANPFLTRDVOHVIVRTSRAGH 391

RESULT 10
US-08-976-838-21
; Sequence 21, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
; MOLECULES
```



```

NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSER: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,838
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 881 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein*
US-08-976-838-21

```

```

Query Match          11.3%; Score 66; DB 2; Length 881;
Best Local Similarity 27.5%; Pred. No. 35;
Matches      25; Conservative    16; Mismatches   26; Indels    24; Gaps     5

QY      37 MSATPKHG----LEOCPPAPRPVVTGFTGDGSAKETVSODKRF---SOGHTWCITLALP-- 86
       :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      303 ISSTAESGKRKPWYLEEC--SSTLATYSSGSSEYDKKITTTDLRORCTDNNHGTASANMA 360
               -----HPMLTW--VGHLRNHVSSASH 105
               :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      361 AGITALALEANPLTWRDVQHVIVRTSRAGH 391

RESULT 11
US-08-525-940-18
; Sequence 18, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; APPLICANT: Wolf, Joseph R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
; TITLE OF INVENTION: ENCODING SAID PROTEASES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

```

```

1 APPLICATION NUMBER: US 08/368,852
2 FILING DATE: 01-JAN-1995
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: US 08/088,322
5 FILING DATE: 07-JUL-1993
6 ATTORNEY/AGENT INFORMATION:
7 NAME: Connell, Gary J.
8 REGISTRATION NUMBER: 32,020
9 REFERENCE/DOCKET NUMBER: 2848-11-C1
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: (303) 863-9700
12 TELEFAX: (303) 863-0223
13 INFORMATION FOR SEQ ID NO: 18:
14 SEQUENCE CHARACTERISTICS:
15 LENGTH: 915 amino acids
16 TYPE: amino acid
17 TOPOLOGY: linear
18 MOLECULE TYPE: protein
19 US-08-525-940-18

```

[illegible]

```

US-08-976-838-18
Sequence 18, Application US/08976838
Patent No. 5981259
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
TITLE OF INVENTION: CD4+ T-LIMPHOCYTE PROTEASE NUCLEIC ACID
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,838
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848 -11-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-976-838-18

```

```

37 MSATPKHG-----LEQCPAPPPAATGFTGDSGAKEVSDKR---SQGHWTCTLALP-- 86
:::| : | | ::|| : ::| | : || : :|

```

```

001      87-----HPWLTV--VGLRNNHVSASH 105
002      :|:|:| | |: | |
003      :|:|:| | |: | |
004      :|:|:| | |: | |
005      :|:|:| | |: | |
006      :|:|:| | |: | |
007      :|:|:| | |: | |
008      :|:|:| | |: | |
009      :|:|:| | |: | |
010      :|:|:| | |: | |
011      :|:|:| | |: | |
012      :|:|:| | |: | |
013      :|:|:| | |: | |
014      :|:|:| | |: | |
015      :|:|:| | |: | |
016      :|:|:| | |: | |
017      :|:|:| | |: | |
018      :|:|:| | |: | |
019      :|:|:| | |: | |
020      :|:|:| | |: | |
021      :|:|:| | |: | |
022      :|:|:| | |: | |
023      :|:|:| | |: | |
024      :|:|:| | |: | |
025      :|:|:| | |: | |
026      :|:|:| | |: | |
027      :|:|:| | |: | |
028      :|:|:| | |: | |
029      :|:|:| | |: | |
030      :|:|:| | |: | |
031      :|:|:| | |: | |
032      :|:|:| | |: | |
033      :|:|:| | |: | |
034      :|:|:| | |: | |
035      :|:|:| | |: | |
036      :|:|:| | |: | |
037      :|:|:| | |: | |
038      :|:|:| | |: | |
039      :|:|:| | |: | |
040      :|:|:| | |: | |
041      :|:|:| | |: | |
042      :|:|:| | |: | |
043      :|:|:| | |: | |
044      :|:|:| | |: | |
045      :|:|:| | |: | |
046      :|:|:| | |: | |
047      :|:|:| | |: | |
048      :|:|:| | |: | |
049      :|:|:| | |: | |
050      :|:|:| | |: | |
051      :|:|:| | |: | |
052      :|:|:| | |: | |
053      :|:|:| | |: | |
054      :|:|:| | |: | |
055      :|:|:| | |: | |
056      :|:|:| | |: | |
057      :|:|:| | |: | |
058      :|:|:| | |: | |
059      :|:|:| | |: | |
060      :|:|:| | |: | |
061      :|:|:| | |: | |
062      :|:|:| | |: | |
063      :|:|:| | |: | |
064      :|:|:| | |: | |
065      :|:|:| | |: | |
066      :|:|:| | |: | |
067      :|:|:| | |: | |
068      :|:|:| | |: | |
069      :|:|:| | |: | |
070      :|:|:| | |: | |
071      :|:|:| | |: | |
072      :|:|:| | |: | |
073      :|:|:| | |: | |
074      :|:|:| | |: | |
075      :|:|:| | |: | |
076      :|:|:| | |: | |
077      :|:|:| | |: | |
078      :|:|:| | |: | |
079      :|:|:| | |: | |
080      :|:|:| | |: | |
081      :|:|:| | |: | |
082      :|:|:| | |: | |
083      :|:|:| | |: | |
084      :|:|:| | |: | |
085      :|:|:| | |: | |
086      :|:|:| | |: | |
087      :|:|:| | |: | |
088      :|:|:| | |: | |
089      :|:|:| | |: | |
090      :|:|:| | |: | |
091      :|:|:| | |: | |
092      :|:|:| | |: | |
093      :|:|:| | |: | |
094      :|:|:| | |: | |
095      :|:|:| | |: | |
096      :|:|:| | |: | |
097      :|:|:| | |: | |
098      :|:|:| | |: | |
099      :|:|:| | |: | |
100      :|:|:| | |: | |

```

RESULT 15

RESULT 15
US-08-053-867A-2
; Sequence 2, Application US/08053867A
; Patent No. 5545545
; GENERAL INFORMATION:
; APPLICANT: Genenbach, Burle G.
; INVENTOR: Genenbach, Burle G.

TITLE OF INVENTION: Lysine-insensitive Maize
 TITLE OF INVENTION: Dihydrodipicolinic Acid Synthase
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merchant & Gould
 STREET: 3100 No. 5545545west Center
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55403
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/053,867A
FILING DATE: 27-Apr-1993
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D.
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 600.263-US-01

TELECOMMUNICATION INFORMATION:
TELEPHONE: 612/332-5300
TELEFAX: 612/332-3081
INFORMATION FOR SEQ ID NO: 2:

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 326 amino acids
: TYPE: amino acid
: TOPOLOGY: 1linear
:

```

[illegible]

Search completed: June 17, 2003, 10:38:13
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:37:01 ; Search time 47 Seconds

(without alignments)
238.854 Million cell updates/sec

Title: US-09-955-807-2

Perfect score: 582

Sequence: 1 MGXSEPMPCAHPLGLFLG.....PHPMLTWGHLRNHVSASH 105

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PTCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	582	100.0	105	10	US-09-955-807-2
2	582	100.0	105	10	US-09-955-807-14
3	443	76.1	80	10	US-09-955-807-16
4	443	76.1	89	10	US-09-955-807-15
5	439	75.4	79	10	US-09-955-807-3
6	428	75.5	77	10	US-09-955-807-4
7	375	64.4	65	10	US-09-955-807-5
8	372	62.9	65	10	US-09-955-807-10
9	246	42.3	43	10	US-09-955-807-12
10	229	39.3	42	10	US-09-955-807-9
11	114	19.6	20	10	US-09-955-807-8
12	103	17.7	20	10	US-09-955-807-11
13	79	13.6	136	9	US-09-764-891-4761
14	76	13.1	380	10	US-09-814-777A-34
15	70	12.0	76	9	US-09-796-692-876
16	70	12.0	76	9	US-10-040-862-876
17	70	12.0	1616	9	US-09-712-363-262
18	69.5	11.9	147	9	US-10-016-634A-115
19	68	11.7	73	9	US-09-796-692-1130

20	68	11.7	73	9	US-10-040-862-1130	Sequence 1130, App
21	67.5	11.6	230	10	US-09-945-676-6	Sequence 6, Appl1
22	67	11.5	253	10	US-09-971-543-2	Sequence 2, Appl1
23	66.5	11.4	179	9	US-09-791-279-120	Sequence 120, App
24	66.5	11.4	180	9	US-09-768-703-2	Sequence 5249, App
25	66	11.3	368	10	US-09-764-891-5249	Sequence 2, Appl1
26	66	11.3	807	9	US-10-063-547-98	Sequence 98, Appl
27	66	11.3	807	9	US-10-063-616-98	Sequence 88, Appl
28	66	11.3	807	9	US-10-063-502-98	Sequence 188, App
29	66	11.3	807	9	US-10-227-884-188	Sequence 188, App
30	66	11.3	807	9	US-10-230-163-188	Sequence 188, App
31	66	11.3	807	9	US-10-006-856A-229	Sequence 229, App
32	66	11.3	807	9	US-10-218-631-188	Sequence 188, App
33	66	11.3	807	9	US-10-230-338-188	Sequence 188, App
34	66	11.3	807	9	US-10-063-518-98	Sequence 98, Appl
35	66	11.3	807	9	US-10-063-598-98	Sequence 98, Appl
36	66	11.3	807	9	US-10-227-693-98	Sequence 98, Appl
37	66	11.3	807	9	US-10-230-414-188	Sequence 188, App
38	66	11.3	807	9	US-10-006-818A-229	Sequence 229, App
39	66	11.3	807	9	US-10-015-393A-229	Sequence 229, App
40	66	11.3	807	9	US-10-063-567-98	Sequence 98, Appl
41	66	11.3	807	9	US-10-216-159A-188	Sequence 188, App
42	66	11.3	807	9	US-09-946-374-229	Sequence 229, App
43	66	11.3	807	9	US-10-012-121A-229	Sequence 229, App
44	66	11.3	807	9	US-10-015-869A-229	Sequence 229, App
45	66	11.3	807	9	US-10-063-538-98	Sequence 98, Appl

ALIGNMENTS

RESULT 1
US-09-955-807-2
Sequence 2, Application US/09955807
Patent No. US20020132996A1
GENERAL INFORMATION:
APPLICANT: Lok, Si
APPLICANT: Sheppard, Paul O.
APPLICANT: Kindsvogel, Wayne
APPLICANT: Bort, Susan J.
TITLE OF INVENTION: Secretory Protein-48
FILE REFERENCE: 98-17C1
CURRENT APPLICATION NUMBER: US/09/955,807
CURRENT FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: 60/102,679
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 09/410,603
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2:
LENGTH: 105
TYPE: PRT
ORGANISM: Homo sapiens
US-09-955-807-2
Query Match 100.0%; Score 582; DB 10; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.5e-51;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGXSEPMPCAHPLGLFLGHPALSLPVTVYAGVMSATPKHGLEQCPAPPAVYGT 60
DB 1 MGXSEPMPCAHPLGLFLGHPALSLPVTVYAGVMSATPKHGLEQCPAPPAVYGT 60
QY 61 GDSGAKETVSODKRSQGTWCTLALPHPWLTWVGHRLNHYSSASH 105
DB 61 GDSGAKETVSODKRSQGTWCTLALPHPWLTWVGHRLNHYSSASH 105
QY 61 GDSGAKETVSODKRSQGTWCTLALPHPWLTWVGHRLNHYSSASH 105
DB 61 GDSGAKETVSODKRSQGTWCTLALPHPWLTWVGHRLNHYSSASH 105
RESULT 2
US-09-955-807-14
Sequence 14, Application US/09955807
Patent No. US20020132996A1

GENERAL INFORMATION:
APPLICANT: Lok, SI
APPLICANT: Sheppard, Paul O.
APPLICANT: Kindsvogel, Wayne
APPLICANT: Bort, Susan J.
TITLE OF INVENTION: Secretory Protein-48
FILE REFERENCE: 98-17C1
CURRENT APPLICATION NUMBER: US/09/955,807
CURRENT FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: 60/102,679
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 09/410,603
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 105
TYPE: PRT
ORGANISM: Homo sapiens
US-09-955-807-14

Query Match
Best Local Similarity 100.0%; Score 582; DB 10; Length 105;
Pred. No. 1.5e-51;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLGSEPMCAHPGLGLGLHPLSLPLVTVAGVMSATPKHGLECCPPAPPAVTGFT 60
DB 1 MLGSEPMCAHPGLGLGLHPLSLPLVTVAGVMSATPKHGLECCPPAPPAVTGFT 60

QY 61 GDSAKETVSODKRSOGHTCTLALPHPWLTWVGHRLRNHVSASH 105
DB 61 GDSAKETVSODKRSOGHTCTLALPHPWLTWVGHRLRNHVSASH 105

RESULT 3
US-09-955-807-16
Sequence 16, Application US/09955807
Patent No. US20020132996A1
GENERAL INFORMATION:
APPLICANT: Lok, SI
APPLICANT: Sheppard, Paul O.
APPLICANT: Kindsvogel, Wayne
APPLICANT: Bort, Susan J.
TITLE OF INVENTION: Secretory Protein-48
FILE REFERENCE: 98-17C1
CURRENT APPLICATION NUMBER: US/09/955,807
CURRENT FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: 60/102,679
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 09/410,603
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 80
TYPE: PRT
ORGANISM: Homo sapiens
US-09-955-807-16

Query Match
Best Local Similarity 100.0%; Score 443; DB 10; Length 80;
Pred. No. 1.1e-37;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 SLPLVTVAGVMSATPKHGLECCPPAPPAVTGFTGDSGAKETVSODKRSOGHTCTLAL 85
DB 1 SLPLVTVAGVMSATPKHGLECCPPAPPAVTGFTGDSGAKETVSODKRSOGHTCTLAL 60

QY 86 PHPMLTWVGHRLRNHVSASH 105
DB 61 PHPMLTWVGHRLRNHVSASH 80

RESULT 4

US-09-955-807-15
Sequence 15, Application US/09955807
Patent No. US20020132996A1
GENERAL INFORMATION:
APPLICANT: Lok, SI
APPLICANT: Sheppard, Paul O.
APPLICANT: Kindsvogel, Wayne
APPLICANT: Bort, Susan J.
TITLE OF INVENTION: Secretory Protein-48
FILE REFERENCE: 98-17C1
CURRENT APPLICATION NUMBER: US/09/955,807
CURRENT FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: 60/102,679
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 09/410,603
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 15
LENGTH: 89
TYPE: PRT
ORGANISM: Homo sapiens
US-09-955-807-15

Query Match
Best Local Similarity 100.0%; Score 443; DB 10; Length 89;
Pred. No. 1.3e-37;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 SLPLVTVAGVMSATPKHGLECCPPAPPAVTGFTGDSGAKETVSODKRSOGHTCTLAL 85
DB 10 SLPLVTVAGVMSATPKHGLECCPPAPPAVTGFTGDSGAKETVSODKRSOGHTCTLAL 69

QY 86 PHPMLTWVGHRLRNHVSASH 105
DB 70 PHPMLTWVGHRLRNHVSASH 89

RESULT 5
US-09-955-807-3
Sequence 3, Application US/09955807
Patent No. US20020132996A1
GENERAL INFORMATION:
APPLICANT: Lok, SI
APPLICANT: Sheppard, Paul O.
APPLICANT: Kindsvogel, Wayne
APPLICANT: Bort, Susan J.
TITLE OF INVENTION: Secretory Protein-48
FILE REFERENCE: 98-17C1
CURRENT APPLICATION NUMBER: US/09/955,807
CURRENT FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: 60/102,679
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 09/410,603
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 79
TYPE: PRT
ORGANISM: Homo sapiens
US-09-955-807-3

Query Match
Best Local Similarity 100.0%; Score 439; DB 10; Length 79;
Pred. No. 2.8e-37;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 LPLVTVAGVMSATPKHGLECCPPAPPAVTGFTGDSGAKETVSODKRSOGHTCTLALP 86
DB 1 LPLVTVAGVMSATPKHGLECCPPAPPAVTGFTGDSGAKETVSODKRSOGHTCTLALP 60

QY 87 HPMLTWVGHRLRNHVSASH 105
DB 61 HPMLTWVGHRLRNHVSASH 79

RESULT 6
US-09-955-807-4
; Sequence 4, Application US/09955807
; Patent No. US20020132996A1
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Bort, Susan J.
; TITLE OF INVENTION: Secretory Protein-48
; FILE REFERENCE: 98-17C1
; CURRENT APPLICATION NUMBER: US/09/955,807
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/102,679
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 09/410,603
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-955-807-4

Query Match
Best Local Similarity 100.0%; Score 428; DB 10; Length 77;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LVTVAGVWATPKIGLECCPPAPPAVTGFTGDSGAKETVSQDKRSQGHWTCTLALPH 88
DB 1 LVTVAGVWATPKIGLECCPPAPPAVTGFTGDSGAKETVSQDKRSQGHWTCTLALPH 60
QY 89 WLTWVGHILRNHVSASH 105
DB 61 WLTWVGHILRNHVSASH 77

RESULT 7
US-09-955-807-5
; Sequence 5, Application US/09955807
; Patent No. US20020132996A1
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Bort, Susan J.
; TITLE OF INVENTION: Secretory Protein-48
; FILE REFERENCE: 98-17C1
; CURRENT APPLICATION NUMBER: US/09/955,807
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/102,679
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 09/410,603
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-955-807-5

Query Match
Best Local Similarity 100.0%; Score 375; DB 10; Length 65;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 SSASH 105
DB 61 SSASH 65

RESULT 8
US-09-955-807-10
; Sequence 10, Application US/09955807
; Patent No. US20020132996A1
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Bort, Susan J.
; TITLE OF INVENTION: Secretory Protein-48
; FILE REFERENCE: 98-17C1
; CURRENT APPLICATION NUMBER: US/09/955,807
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/102,679
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 09/410,603
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-955-807-10

Query Match
Best Local Similarity 100.0%; Score 372; DB 10; Length 65;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 SATPKHGLECCPPAPPAVTGFTGDSGAKETVSQDKRSQGHWTCTLALPHMLTWVGHIL 97
DB 1 SATPKHGLECCPPAPPAVTGFTGDSGAKETVSQDKRSQGHWTCTLALPHMLTWVGHIL 60
QY 98 NHVSS 102
DB 61 NHVSS 65

RESULT 9
US-09-955-807-12
; Sequence 12, Application US/09955807
; Patent No. US20020132996A1
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Bort, Susan J.
; TITLE OF INVENTION: Secretory Protein-48
; FILE REFERENCE: 98-17C1
; CURRENT APPLICATION NUMBER: US/09/955,807
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/102,679
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 09/410,603
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-955-807-12

Query Match
Best Local Similarity 100.0%; Score 246; DB 10; Length 43;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 TGDGSAKETVSQDKRSQGHWTCTLALPHMLTWVGHILRNHVS 102

Db 1 TGDGSAKETVSQDKRSQGHWTCTALPMPWLTWGHILKNHVS 43

RESULT 10

US-09-955-807-9
; Sequence 9, Application US/09955807
; Patent No. US20020132996A1
; GENERAL INFORMATION:
; APPLICANT: Lok, S.I.
; APPLICANT: Shepard, Paul O.
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Bort, Susan J.
; TITLE OF INVENTION: Secretory Protein-48
; FILE REFERENCE: 98-17C1
; CURRENT APPLICATION NUMBER: US/09/955,807
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/102,679
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 09/410,603
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-955-807-9

Query Match

Best Local Similarity 100.0%; Score 229; DB 10; Length 42;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 SATPKHGLEQCPAPPAVYTGFTGDSGAKETVSQDKRSQGH 79
Db 1 SATPKHGLEQCPAPPAVYTGFTGDSGAKETVSQDKRSQGH 42

RESULT 11

US-09-955-807-8
; Sequence 8, Application US/09955807
; Patent No. US20020132996A1
; GENERAL INFORMATION:
; APPLICANT: Lok, S.I.
; APPLICANT: Shepard, Paul O.
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Bort, Susan J.
; TITLE OF INVENTION: Secretory Protein-48
; FILE REFERENCE: 98-17C1
; CURRENT APPLICATION NUMBER: US/09/955,807
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/102,679
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 09/410,603
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-955-807-8

Query Match

Best Local Similarity 100.0%; Score 114; DB 10; Length 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 SATPKHGLEQCPAPPAVY 57
Db 1 SATPKHGLEQCPAPPAVY 20

RESULT 12

US-09-955-807-11
; Sequence 11, Application US/09955807
; Patent No. US20020132996A1
; GENERAL INFORMATION:
; APPLICANT: Lok, S.I.
; APPLICANT: Shepard, Paul O.
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Bort, Susan J.
; TITLE OF INVENTION: Secretory Protein-48
; FILE REFERENCE: 98-17C1
; CURRENT APPLICATION NUMBER: US/09/955,807
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/102,679
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 09/410,603
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-955-807-11

Query Match

Best Local Similarity 100.0%; Score 103; DB 10; Length 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 TGDGSAKETVSQDKRSQGH 79
Db 1 TGDGSAKETVSQDKRSQGH 20

RESULT 13

US-09-764-891-4761
; Sequence 4761, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4761
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (25)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (42)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4761

Query Match

Best Local Similarity 13.6%; Score 79; DB 9; Length 136;
Matches 33; Conservative 5; Mismatches 36; Indels 42; Gaps 5;

OY 7 PRCANHLGLFLGLHRLSLPLVYVYAGVMSAT-----PKHGLEQCPAPPAVYTGFTG 61
Db 33 PRCANHLGLFLGLHRLSLPLVYVYAGVMSAT-----PKHGLEQCPAPPAVYTGFTG 78
OY 62 DSGAKETVSQDKRSQGH-----TWCTLALPMPWLTWGH 94
Db 79 ESNPYPILTSQVSYSHMDVYMRHPLMPTCTTATATITTTTLALPMPWLTWGH 134

```

RESULT 14
US-09-814-777A-34
: Sequence 34, Application US/09814777A
: Patent No. US20020142415A1
: GENERAL INFORMATION:
: APPLICANT: KOOPMAN, Peter Anthony
: APPLICANT: MUSCAT, George Eugene Orlando
: TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES AND METHODS OF USING THEM
: FILE REFERENCE: 21415-0003
: CURRENT APPLICATION NUMBER: US/09/814,777A
: PRIORITY FILING DATE: 2001-03-23
: PRIOR APPLICATION NUMBER: AU PQ6457
: PRIOR FILING DATE: 2000-03-24
: NUMBER OF SEQ ID NOS: 128
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 34
: LENGTH: 380
: TYPE: PRF
: ORGANISM: Mouse
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (180)..(392)
: OTHER INFORMATION: HMG; Region: high mobility group
: NAME/KEY: misc_feature
: LOCATION: (183)..(389)
: OTHER INFORMATION: HMG box
US-09-814-777A-34

```

```

Query March 13.1% Score 76; DB 10; Length 380;
Best Local Similarity 24.0%; Pred. No. 7;
Matches 31; Conservative 9; Mismatches 31; Indels 58; Gaps 6;

QY 7 PMPCANPLGLFLGLALHPALSLPLVTVAGVMSATPKHGLECCP----- 50
   | : | | | | | | | | | : | : | : | | |
Db 242 PLHCSEHPLGLSALGQSPGV--MMSVSG-----CPESPAYVSHATYHPLPN 287
   | : | | | | | | | | | : | | | | | |

QY 51 -----APPVATG-----TQDSGAKETYSQDKRSQGTWCLALPHRMLT 91
   | | | | | | | | | | | | | | | | | |

Db 288 LQAHLGQISPEHPHGFDTLDQLSQVELLSGDMRNE-FDQYLNTPGHPDSAA----- 339
   | | | | | | | | | | | | | | | | | |

QY 92 VVGHILRNHV 100
   | | | | | | | |

Db 340 -VGTLDGTHV 347

RESULT 15
US-09-796-692-876
; Sequence 876, Application US/09796692
; Publication No. US20020198362A1
GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084

```

```

? PRIOR FILING DATE: 2000-05-04
? PRIOR APPLICATION NUMBER: 60/206, 201
? PRIOR FILING DATE: 2000-05-22
? PRIOR APPLICATION NUMBER: 60/218, 950
? PRIOR FILING DATE: 2000-07-14
? PRIOR APPLICATION NUMBER: 60/222, 903
? PRIOR FILING DATE: 2000-08-03
? PRIOR APPLICATION NUMBER: 60/223, 416
? PRIOR FILING DATE: 2000-08-04
? PRIOR APPLICATION NUMBER: 60/223, 378
? PRIOR FILING DATE: 2000-08-07
? NUMBER OF SEQ ID NOS: 9597
? SOFTWARE: FastSeq for Windows Version 3.0.
? SEQ ID NO 876
? LENGTH: 76
? TYPE: prt
? ORGANISM: Homo sapiens
US-03-796-692-876

```

QY	43	HGLEQCPAPPANTGFTG	-----DSGAKETVSSQDKRSQGHV	-----	CTLAIP	86
Db	9	HGLRSPRPPLSLHRFCDFLCVSDSLTSLVSGPRRGGDGLTMGKQWESQCTLLVP				64

Query Match 12.0%; Score 70; DB 9; Length 76;
 Best Local Similarity 32.1%; Pred. No. 4.4;
 Matches 18; Conservative 5; Mismatches 21; Indels 12; Gaps 2.

Search completed: June 17, 2003, 10:46:19
Job time : 49 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:25:05 ; Search time 70 Seconds

(without alignments)
199.876 Million cell updates/sec

Title: US-09-955-807-2

Perfect score: 582

Sequence: 1 MGXSEMPACAHPLGLFLLG.....PMPMLWVGHILRNHVSASH 105

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
A.Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	582	100.0	105	21	AAV79436 Human secretory pr
2	443	76.1	80	21	AAV79440 Human secretory pr
3	443	76.1	89	21	AAV79446 Human secretory pr
4	439	75.4	79	21	AAV79437 Human secretory pr
5	428	73.5	77	21	AAV79438 Human secretory pr
6	375	64.4	65	21	AAV79439 Human secretory pr
7	372	63.9	65	21	AAV79443 Human secretory pr
8	246	42.3	43	21	AAV79445 Human secretory pr
9	229	39.3	42	21	AAV79442 Human secretory pr
10	114	19.6	20	21	AAV79441 Human secretory pr

11	103	17.7	20	21	AAV79444 Human secretory pr
12	89	15.3	743	22	ABB68417 Drosophila melanog
13	79.5	13.7	252	21	ABB42156 Human ORFX ORF1920
14	79.5	13.7	598	23	ABB97893 Human secretory po
15	79	13.6	136	22	AAV61033 Human reproductive
16	79	13.6	188	21	AAV38275 Human secreted pro
17	79	13.6	188	21	AAV38277 Gene 19 human secr
18	77	13.2	172	21	AAV41678 Human ORFX ORF1442
19	75.5	13.0	75	22	AAV49832 Propionibacterium
20	73.5	12.6	138	20	AAV73090 Amino acid sequenc
21	73.5	12.6	941	23	ABB08868 Cricetus griseus
22	72.5	12.5	364	21	AAV53186 Macaca mulatta rha
23	72	12.4	141	22	AAV79666 Human protein seq
24	72	12.4	572	22	AAV32685 Novel human secret
25	71.5	12.3	146	22	AAV61349 Propionibacterium
26	71.5	12.3	989	22	AAV90769 Human shear stress
27	71	12.2	93	22	ABG19572 Novel human diago
28	71	12.2	415	22	ABG09627 Novel human diago
29	71	12.2	900	21	AAV42321 Human ORFX ORF2085
30	71	12.2	1217	22	ABG09876 Novel human diago
31	71	12.2	1239	22	ABG09857 Mus musculus notch
32	71	12.2	1964	20	AAV95557 Human Persephin-AR
33	70.5	12.1	164	21	AAV44836 Novel human diago
34	70.5	12.1	813	22	ABG25423 Human haematologic
35	70	12.0	76	22	AAV80512 Mycobacterium tube
36	70	12.0	1615	22	AAV81211 Human EST encoded
37	69.5	11.9	186	22	AAV24399 Propionibacterium
38	69.5	11.9	305	22	AAV58771 Novel human diago
39	69	11.9	461	22	AAV03202 Amino acid sequenc
40	69	11.9	423	21	AAV10280 Murine adult thymu
41	69	11.9	447	22	ABB66646 Drosophila melanog
42	69	11.9	137	23	ABB89082 Human polypeptide
43	68.5	11.8	137	23	AAV42298 Propionibacterium
44	68.5	11.8	275	22	AAV42100 Human polypeptide
45	68.5	11.8	301	22	AAV42100 Human polypeptide

ALIGNMENTS

RESULT 1	AAV79436	standard; Protein; 105 AA.
ID	AAV79436	
XX	AAV79436;	
AC	01-AUG-2000	(first entry)
DT		
XX		
DE		Human secretory protein 48 (Zsig48).
XX		
KW		Secretory protein 48; Zsig48; human; leukocyte; proliferation;
KW		immunostimulant; adjuvant; therapy.
XX		
OS		Homo sapiens.
XX		
XX		
FT	Key	Location/Qualifiers
FT	Peptide	1..26
FT		/note= "signal peptide; alternatively comprises residues 1-25, 1-28 or 1-40"
FT	Protein	27..105
FT		/note= "mature protein; alternative comprises residues 26-105, 29-105 or 41-105"
FT	Peptide	38..57
FT		/note= "antigen peptide; a polypeptide containing this sequence is specifically claimed in Claim 2"
FT	Peptide	38..79
FT		/note= "antigen peptide; a polypeptide containing this sequence is specifically claimed in Claim 2"
FT	Peptide	38..102
FT		/note= "antigen peptide; a polypeptide containing this sequence is specifically claimed in Claim 2"

FT Peptide 60..79 Claim 2"
 FT /note- "antigen peptide; a polypeptide containing
 FT this sequence is specifically claimed in
 FT Claim 2"
 FT Peptide 60..102
 FT /note- "antigen peptide; a polypeptide containing
 FT this sequence is specifically claimed in
 FT Claim 2"
 XX
 PN WO200018796-A2.
 XX
 PD 06-APR-2000.
 XX
 PF 01-OCT-1999; 99WO-US22970.
 XX
 PR 01-OCT-1998; 98US-0164740.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Lok S, Sheppard PO;
 XX
 DR WPI: 2000-303441/26.
 DR N-PSDB: AA294889, AA284892.
 XX
 PS Claim 1; Page 128-129; 141pp; English.
 XX
 CC The present sequence is that of human secretory protein 48 (Zs1g48),
 CC a protein that stimulates the proliferation of peripheral blood
 CC mononuclear cells, i.e., T-cells, B-cells and monocytes. The
 CC invention provides Zs1g48 polynucleotides, polypeptides, antibodies
 CC and anti-idiotypic antibodies. Zs1g48 stimulates the proliferation
 CC of leukocytes in both a mixed leukocyte reaction and in an unmixe
 CC leukocyte reaction. It can be used to promote proliferation of
 CC peripheral blood leukocytes and to treat low leukocyte counts in
 CC individuals (claimed). This can be useful in treating cancer
 CC patients whose leukocytes have been depleted by chemotherapy,
 CC radiation or illness. Zs1g48 can be administered to patients
 CC receiving bone marrow transplants to promote proliferation of
 CC leukocytes produced by the transplanted marrow. It could also be
 CC useful in treating immunosuppressed individuals such as the elderly
 CC or HIV infected individuals, or used as a vaccine adjuvant.
 CC
 XX
 SQ Sequence 105 AA;
 Query Match 100.0%; Score 582; DB 21; Length 105;
 Best Local Similarity 100.0%; Pred. No. 2e-54;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLCGSEPMCAHPIGLGLSLHPLSLPLVYTVAGVMSATPRKHGLECCPPAPPAVYGT 60
 DB 1 MLCGSEPMCAHPIGLGLSLHPLSLPLVYTVAGVMSATPRKHGLECCPPAPPAVYGT 60
 QY 61 GDSGAKETVSODKRSOGHTWCTTALPHPLTWVGHILNNHVSASH 105
 DB 61 GDSGAKETVSODKRSOGHTWCTTALPHPLTWVGHILNNHVSASH 105
 RESULT 2
 ID AAY79440 standard; Protein: 80 AA.
 AC AAY79440;
 DT 01-AUG-2000 (first entry)
 XX
 DE Human secretory protein 48 (Zs1g48) mature polypeptide.
 XX
 KW Secretory protein 48; Zs1g48; human; leukocyte; proliferation;

KW Immunostimulant; adjuvant; therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200018796-A2.
 XX
 PD 06-APR-2000.
 XX
 PF 01-OCT-1999; 99WO-US22970.
 XX
 PR 01-OCT-1998; 98US-0164740.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Lok S, Sheppard PO;
 XX
 DR WPI: 2000-303441/26.
 DR
 XX
 PS Claim 1; Page 138-139; 141pp; English.
 XX
 CC The present sequence is that of human secretory protein 48 (Zs1g48)
 CC mature polypeptide, i.e., lacking the signal sequence (see AAY79436).
 CC Zs1g48 mature protein stimulates the proliferation of peripheral
 CC blood mononuclear cells, i.e., T-cells, B-cells and monocytes. The
 CC invention provides Zs1g48 polynucleotides, polypeptides, antibodies
 CC and anti-idiotypic antibodies. Zs1g48 stimulates the proliferation
 CC of leukocytes in both a mixed leukocyte reaction and in an unmixe
 CC leukocyte reaction. It can be used to promote proliferation of
 CC peripheral blood leukocytes and to treat low leukocyte counts in
 CC individuals (claimed). This can be useful in treating cancer
 CC patients whose leukocytes have been depleted by chemotherapy,
 CC radiation or illness. Zs1g48 can be administered to patients
 CC receiving bone marrow transplants to promote proliferation of
 CC leukocytes produced by the transplanted marrow. It could also be
 CC useful in treating immunosuppressed individuals such as the elderly
 CC or HIV infected individuals, or used as a vaccine adjuvant.
 CC
 XX
 SQ Sequence 80 AA;
 Query Match 76.1%; Score 443; DB 21; Length 80;
 Best Local Similarity 100.0%; Pred. No. 9.4e-40;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 26 SLPLVTVAGVMSATPRKHGLECCPPAPPAVYGTGDSGAKETVSODKRSOGHTWCTTAL 85
 DB 1 SLPLVTVAGVMSATPRKHGLECCPPAPPAVYGTGDSGAKETVSODKRSOGHTWCTTAL 60
 QY 86 PHPMLTWVGHILNNHVSASH 105
 DB 61 PHPMLTWVGHILNNHVSASH 80
 RESULT 3
 ID AAY79446 standard; Protein: 89 AA.
 AC AAY79446;
 DT 01-AUG-2000 (first entry)
 XX
 DE Human secretory protein 48 (Zs1g48) fusion protein.
 XX
 KW Secretory protein 48; Zs1g48; human; leukocyte; proliferation;
 KW Immunostimulant; adjuvant; therapy; fusion protein.
 XX
 OS Homo sapiens.
 XX
 DE Key Location/Qualifiers
 FT Protein 10..89

FT /note= "Zsig48 mature polypeptide"
 XX
 PN W0200018796-A2.
 XX
 PD 06-APR-2000.
 XX
 PF 01-OCT-1999; 99WO-US22970.
 XX
 PR 01-OCT-1998; 98US-0164740.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Lok S, Sheppard PO;
 XX
 DR WPI; 2000-303441/26.
 XX
 PT Human secretory protein 48 polypeptides and polynucleotides useful for
 PT promoting leukocyte proliferation and for treating immunosuppressed
 PT individuals -
 XX
 PS Claim 1; Page 138; 141pp; English.
 XX
 CC The present sequence is that of a human secretory protein 48
 CC (Zsig48) fusion protein. Zsig48 stimulates the proliferation of
 CC peripheral blood mononuclear cells, i.e. T-cells, B-cells and
 CC monocytes. The invention provides Zsig48 polynucleotides,
 CC polypeptides, fusion proteins, antibodies and anti-idiotypic
 CC antibodies. Zsig48 stimulates the proliferation of leukocytes in
 CC both a mixed leukocyte reaction and in an unmixd leukocyte
 CC reaction. It can be used to promote proliferation of peripheral
 CC blood leukocytes and to treat low leukocyte counts in individuals
 CC (claimed). This can be useful in treating cancer patients whose
 CC leukocytes have been depleted by chemotherapy, radiation or
 CC illness. Zsig48 can be administered to patients receiving bone
 CC marrow transplants to promote proliferation of leukocytes produced
 CC by the transplanted marrow. It could also be useful in treating
 CC immunosuppressed individuals such as the elderly or HIV infected
 CC individuals, or used as a vaccine adjuvant.
 XX
 SQ Sequence 89 AA:
 Query Match 76.1%; Score 443; DB 21; Length 89;
 Best Local Similarity 100.0%; Pred. No. 1.1e-39;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 26 SLPLVTVAGVMSATPKHGLEQCPAPPAPVATGFTGDSGAKETVSQDKRSQGHWTCTLAL 85
 DB 10 SLPLVTVAGVMSATPKHGLEQCPAPPAPVATGFTGDSGAKETVSQDKRSQGHWTCTLAL 69
 QY 86 PPHPLTWGHLRNHVSSASH 105
 DB 70 PPHPLTWGHLRNHVSSASH 89
 RESULT 4
 AAY79437
 ID AAY79437 standard; Protein: 79 AA.
 XX
 AC AAY79437;
 XX
 DT 01-AUG-2000 (first entry)
 XX
 DE Human secretory protein 48 (Zsig48) mature polypeptide.
 XX
 KW Secretory protein 48; Zsig48; human; leukocyte; proliferation;
 KW immunostimulant; adjuvant; therapy.
 XX
 OS Homo sapiens.
 XX
 PN W0200018796-A2.
 XX
 PD 06-APR-2000.

PF 01-OCT-1999; 99WO-US22970.
 XX
 PR 01-OCT-1998; 98US-0164740.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Lok S, Sheppard PO;
 XX
 DR WPI; 2000-303441/26.
 XX
 PT Human secretory protein 48 polypeptides and polynucleotides useful for
 PT promoting leukocyte proliferation and for treating immunosuppressed
 PT individuals -
 XX
 PS Claim 1; Page 129; 141pp; English.
 XX
 CC The present sequence is that of human secretory protein 48 (Zsig48)
 CC mature polypeptide, i.e. lacking the signal sequence (see AAY79436).
 CC Zsig48 mature protein stimulates the proliferation of peripheral
 CC blood mononuclear cells, i.e. T-cells, B-cells and monocytes. The
 CC invention provides Zsig48 polynucleotides, polypeptides, antibodies
 CC and anti-idiotypic antibodies. Zsig48 stimulates the proliferation
 CC of leukocytes in both a mixed leukocyte reaction and in an unmixd
 CC leukocyte reaction. It can be used to promote proliferation of
 CC peripheral blood leukocytes and to treat low leukocyte counts in
 CC individuals (claimed). This can be useful in treating cancer
 CC patients whose leukocytes have been depleted by chemotherapy,
 CC radiation or illness. Zsig48 can be administered to patients
 CC receiving bone marrow transplants to promote proliferation of
 CC leukocytes produced by the transplanted marrow. It could also be
 CC useful in treating immunosuppressed individuals such as the elderly
 CC or HIV infected individuals, or used as a vaccine adjuvant.
 XX
 SQ Sequence 79 AA:
 Query Match 75.4%; Score 439; DB 21; Length 79;
 Best Local Similarity 100.0%; Pred. No. 2.5e-39;
 Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 27 LPLVTVAGVMSATPKHGLEQCPAPPAPVATGFTGDSGAKETVSQDKRSQGHWTCTLALP 86
 DB 1 LPLVTVAGVMSATPKHGLEQCPAPPAPVATGFTGDSGAKETVSQDKRSQGHWTCTLALP 60
 QY 87 HPHPLTWGHLRNHVSSASH 105
 DB 61 HPHPLTWGHLRNHVSSASH 79
 RESULT 5
 AAY79438
 ID AAY79438 standard; Protein: 77 AA.
 XX
 AC AAY79438;
 XX
 DT 01-AUG-2000 (first entry)
 XX
 DE Human secretory protein 48 (Zsig48) mature polypeptide.
 XX
 KW Secretory protein 48; Zsig48; human; leukocyte; proliferation;
 KW immunostimulant; adjuvant; therapy.
 XX
 OS Homo sapiens.
 XX
 PN W0200018796-A2.
 XX
 PD 06-APR-2000.
 XX
 PF 01-OCT-1999; 99WO-US22970.
 XX
 PR 01-OCT-1998; 98US-0164740.
 XX
 PA (ZYMO) ZYMOGENETICS INC.

PI Lok S, Sheppard PO;
XX
XX WPI; 2000-303441/26.
XX
XX Human secretory protein 48 polypeptides and polynucleotides useful for
PT promoting leukocyte proliferation and for treating immunosuppressed
PT individuals -
XX
PS Claim 1; Page 129; 141pp; English.
XX
CC The present sequence is that of human secretory protein 48 (Zs1g48)
CC mature polypeptide, i.e. lacking the signal sequence (see AAY79436).
CC Zs1g48 mature protein stimulates the proliferation of peripheral
CC blood mononuclear cells, i.e. T-cells, B-cells and monocytes. The
CC invention provides Zs1g48 polynucleotides, polypeptides, antibodies
CC and anti-idiotypic antibodies. Zs1g48 stimulates the proliferation
CC of leukocytes in both a mixed leukocyte reaction and in an unmixed
CC leukocyte reaction. It can be used to promote proliferation of
CC peripheral blood leukocytes and to treat low leukocyte counts in
CC individuals (claimed). This can be useful in treating cancer
CC patients whose leukocytes have been depleted by chemotherapy,
CC radiation or illness. Zs1g48 can be administered to patients
CC receiving bone marrow transplants to promote proliferation of
CC leukocytes produced by the transplanted marrow. It could also be
CC useful in treating immunosuppressed individuals such as the elderly
CC or HIV infected individuals, or used as a vaccine adjuvant.
XX
SQ Sequence 77 AA;
Query Match 73.5%; Score 428; DB 21; Length 77;
Best Local Similarity 100.0%; Pred. No. 3.6e-38;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 29 LVVTAVGVSATPKHGECPPAPPAVGTGFTGDSGAKETVSQDKRSQGTWCTIALP
DB 1 LVVTAVGVSATPKHGECPPAPPAVGTGFTGDSGAKETVSQDKRSQGTWCTIALP
QY 89 WLTWVGHLRNHVSASH 105
DB 61 WLTWVGHLRNHVSASH 77
RESULT 6
AAY79439
ID AAY79439 standard; Protein; 65 AA.
XX
AC AAY79439;
XX
DT 01-AUG-2000 (first entry)
XX
DE Human secretory protein 48 (Zs1g48) mature polypeptide.
XX
KM Secretory protein 48; Zs1g48; human; leukocyte; proliferation;
KM immunostimulant; adjuvant; therapy.
XX
OS Homo sapiens.
XX
PN WO200018796-A2.
PD
XX
PD 06-APR-2000.
XX
PF 01-OCT-1999; 99WO-US22970.
XX
PR 01-OCT-1998; 98US-0164740.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Lok S, Sheppard PO;
XX
XX WPI; 2000-303441/26.
PT Human secretory protein 48 polypeptides and polynucleotides useful for
PT promoting leukocyte proliferation and for treating immunosuppressed

PT individuals -
XX
XX Claim 1; Page 130; 141pp; English.
XX
CC The present sequence is that of human secretory protein 48 (Zs1g48)
CC mature polypeptide, i.e. lacking the signal sequence (see AAY79436).
CC Zs1g48 mature protein stimulates the proliferation of peripheral
CC blood mononuclear cells, i.e. T-cells, B-cells and monocytes. The
CC invention provides Zs1g48 polynucleotides, polypeptides, antibodies
CC and anti-idiotypic antibodies. Zs1g48 stimulates the proliferation
CC of leukocytes in both a mixed leukocyte reaction and in an unmixed
CC leukocyte reaction. It can be used to promote proliferation of
CC peripheral blood leukocytes and to treat low leukocyte counts in
CC individuals (claimed). This can be useful in treating cancer
CC patients whose leukocytes have been depleted by chemotherapy,
CC radiation or illness. Zs1g48 can be administered to patients
CC receiving bone marrow transplants to promote proliferation of
CC leukocytes produced by the transplanted marrow. It could also be
CC useful in treating immunosuppressed individuals such as the elderly
CC or HIV infected individuals, or used as a vaccine adjuvant.
XX
SQ Sequence 65 AA;
Query Match 64.4%; Score 375; DB 21; Length 65;
Best Local Similarity 100.0%; Pred. No. 1.3e-32;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 PKHGECPPAPPAVGTGFTGDSGAKETVSQDKRSQGTWCTIALP
DB 1 PKHGECPPAPPAVGTGFTGDSGAKETVSQDKRSQGTWCTIALP
QY 101 SSASH 105
DB 61 SSASH 65
RESULT 7
AAY79443
ID AAY79443 standard; Peptide; 65 AA.
XX
AC AAY79443;
XX
DT 01-AUG-2000 (first entry)
XX
DE Human secretory protein 48 (Zs1g48) antigenic peptide.
XX
KM Secretory protein 48; Zs1g48; human; leukocyte; proliferation;
KM immunostimulant; adjuvant; therapy; antigen; immunogen; antibody.
XX
OS Homo sapiens.
XX
PN WO200018796-A2.
PD
XX
PD 06-APR-2000.
XX
PF 01-OCT-1999; 99WO-US22970.
XX
PR 01-OCT-1998; 98US-0164740.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Lok S, Sheppard PO;
XX
XX WPI; 2000-303441/26.
PT Human secretory protein 48 polypeptides and polynucleotides useful for
PT promoting leukocyte proliferation and for treating immunosuppressed
PT individuals -
XX
PS Claim 2; Page 131; 141pp; English.
XX
CC The present sequence is that of an antigenic peptide of human
CC secretory protein 48 (Zs1g48, see also AAY79436-40), a protein that

CC stimulates the proliferation of peripheral blood mononuclear cells,
 CC i.e. T-cells, B-cells and monocytes. The invention provides Zs1g48
 CC polynucleotides, polypeptides, antibodies and anti-idiotypic
 CC antibodies. Zs1g48 stimulates the proliferation of leukocytes in
 CC both a mixed leukocyte reaction and in an unimixed leukocyte
 CC reaction. It can be used to promote proliferation of peripheral
 CC blood leukocytes and to treat low leukocyte counts in individuals
 CC (claimed). This can be useful in treating cancer patients whose
 CC leukocytes have been depleted by chemotherapy, radiation or illness.
 CC Zs1g48 can be administered to patients receiving bone marrow
 CC transplants to promote proliferation of leukocytes produced by the
 CC transplanted marrow. It could also be useful in treating
 CC immunosuppressed individuals such as the elderly or HIV infected
 CC individuals, or used as a vaccine adjuvant. Anti-Zs1g48 antibodies
 CC can be produced using antigenic Zs1g48 epitope-bearing peptides.
 CC such as the present sequence and used to detect Zs1g48 polypeptides.

XX Sequence 65 AA;

Query Match 63.9%; Score 372; DB 21; Length 65;
 Best Local Similarity 100.0%; Pred. No. 2.7e-32;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 SATPRKHGEQCPAPPAPVATGFTGDSGAKETVSODKRSQGHWTCTLALPHPWLTWVGHRL 97
 DB 1 SATPRKHGEQCPAPPAPVATGFTGDSGAKETVSODKRSQGHWTCTLALPHPWLTWVGHRL 60
 OY 98 NHVSS 102
 DB 61 NHVSS 65

RESULT 8
 AAY79445

ID AAY79445 standard; Peptide: 43 AA.

AC AAY79445;

DT 01-AUG-2000 (first entry)

DE Human secretory protein 48 (Zs1g48) antigenic peptide.

XX Secretory protein 48; Zs1g48; human; leukocyte; proliferation;
 KW immunostimulant; adjuvant; therapy; antigen; immunogen; antibody.

OS Homo sapiens.

PN WO200018796-A2.

PD 06-APR-2000.

PE 01-OCT-1999; 99WO-US22970.

PR 01-OCT-1998; 98US-0164740.

PA (ZYMO) ZYMOGENETICS INC.

PI Lok S, Shepard PO;

DR WPI; 2000-303441/26.

PT Human secretory protein 48 polypeptides and polynucleotides useful for
 PT promoting leukocyte proliferation and for treating immunosuppressed
 PT individuals -

PS Claim 2; Page 132; 141pp; English.

CC The present sequence is that of an antigenic peptide of human
 CC secretory protein 48 (Zs1g48, see also AAY79436-40), a protein that
 CC stimulates the proliferation of peripheral blood mononuclear cells,
 CC i.e. T-cells, B-cells and monocytes. The invention provides Zs1g48
 CC polynucleotides, polypeptides, antibodies and anti-idiotypic
 CC antibodies. Zs1g48 stimulates the proliferation of leukocytes in

CC both a mixed leukocyte reaction and in an unimixed leukocyte
 CC reaction. It can be used to promote proliferation of peripheral
 CC blood leukocytes and to treat low leukocyte counts in individuals
 CC (claimed). This can be useful in treating cancer patients whose
 CC leukocytes have been depleted by chemotherapy, radiation or illness.
 CC Zs1g48 can be administered to patients receiving bone marrow
 CC transplants to promote proliferation of leukocytes produced by the
 CC transplanted marrow. It could also be useful in treating
 CC immunosuppressed individuals such as the elderly or HIV infected
 CC individuals, or used as a vaccine adjuvant. Anti-Zs1g48 antibodies
 CC can be produced using antigenic Zs1g48 epitope-bearing peptides.
 CC such as the present sequence and used to detect Zs1g48 polypeptides.

XX Sequence 43 AA;

Query Match 42.3%; Score 246; DB 21; Length 43;
 Best Local Similarity 100.0%; Pred. No. 4.4e-19;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 TGDGSAKETVSODKRSQGHWTCTLALPHPWLTWVGHRLNHVSS 102
 DB 1 TGDGSAKETVSODKRSQGHWTCTLALPHPWLTWVGHRLNHVSS 43

RESULT 9
 AAY79442

ID AAY79442 standard; Peptide: 42 AA.

AC AAY79442;

DT 01-AUG-2000 (first entry)

DE Human secretory protein 48 (Zs1g48) antigenic peptide.

XX Secretory protein 48; Zs1g48; human; leukocyte; proliferation;
 KW immunostimulant; adjuvant; therapy; antigen; immunogen; antibody.

OS Homo sapiens.

PN WO200018796-A2.

PD 06-APR-2000.

PE 01-OCT-1999; 99WO-US22970.

PR 01-OCT-1998; 98US-0164740.

PA (ZYMO) ZYMOGENETICS INC.

PI Lok S, Shepard PO;

DR WPI; 2000-303441/26.

PT Human secretory protein 48 polypeptides and polynucleotides useful for
 PT promoting leukocyte proliferation and for treating immunosuppressed
 PT individuals -

PS Claim 2; Page 131; 141pp; English.

CC The present sequence is that of an antigenic peptide of human
 CC secretory protein 48 (Zs1g48, see also AAY79436-40), a protein that
 CC stimulates the proliferation of peripheral blood mononuclear cells,
 CC i.e. T-cells, B-cells and monocytes. The invention provides Zs1g48
 CC polynucleotides, polypeptides, antibodies and anti-idiotypic
 CC antibodies. Zs1g48 stimulates the proliferation of leukocytes in
 CC both a mixed leukocyte reaction and in an unimixed leukocyte
 CC reaction. It can be used to promote proliferation of peripheral
 CC blood leukocytes and to treat low leukocyte counts in individuals
 CC (claimed). This can be useful in treating cancer patients whose
 CC leukocytes have been depleted by chemotherapy, radiation or illness.
 CC Zs1g48 can be administered to patients receiving bone marrow
 CC transplants to promote proliferation of leukocytes produced by the
 CC transplanted marrow. It could also be useful in treating

CC Immunosuppressed individuals such as the elderly or HIV infected
CC individuals, or used as a vaccine adjuvant. Anti-Zs1g48 antibodies
CC can be produced using antigenic Zs1g48 epitope-bearing peptides
CC such as the present sequence and used to detect Zs1g48 polypeptides.
XX
SQ Sequence 42 AA;

Query Match 39.3%; Score 229; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 2.8e-17;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 SATPKHGLECCPPAPPAVTGFTGDSGAKETVSODKRSQCHT 79
DB 1 SATPKHGLECCPPAPPAVTGFTGDSGAKETVSODKRSQCHT 42

RESULT 10
AAV79441
ID AAV79441 standard; Peptide: 20 AA.
AC AAV79441;

DT 01-AUG-2000 (first entry)
XX

DE Human secretory protein 48 (Zs1g48) antigenic peptide.

KW Secretory protein 48; Zs1g48; human; leukocyte; proliferation;
KM Immunostimulant; adjuvant; therapy; antigen; immunogen; antibody.
XX

OS Homo sapiens.

PN W0200018796-A2.

PD 06-APR-2000.

PF 01-OCT-1999; 99MO-US22970.

PR 01-OCT-1998; 98US-0164740.

PA (ZYMO) ZYMOGENETICS INC.

PI Lok S, Shepard PO;

DR WPI; 2000-303441/26.

PT Human secretory protein 48 polypeptides and polynucleotides useful for
PT promoting leukocyte proliferation and for treating immunosuppressed
PT individuals -
XX
XX

PS Claim 2; Page 131; 141pp; English.

XX The present sequence is that of an antigenic peptide of human
CC secretory protein 48 (Zs1g48, see also AAV79436-40), a protein that
CC stimulates the proliferation of peripheral blood mononuclear cells,
CC i.e. T-cells, B-cells and monocytes. The invention provides Zs1g48
CC polynucleotides, polypeptides, antibodies and anti-idiotypic
CC antibodies. Zs1g48 stimulates the proliferation of leukocytes in
CC both a mixed leukocyte reaction and in an unmixd leukocyte
CC reaction. It can be used to promote proliferation of peripheral
CC blood leukocytes and to treat low leukocyte counts in individuals
CC (claimed). This can be useful in treating cancer patients whose
CC leukocytes have been depleted by chemotherapy, radiation or illness.
CC Zs1g48 can be administered to patients receiving bone marrow
CC transplants to promote proliferation of leukocytes produced by the
CC immunosuppressed individuals such as the elderly or HIV infected
CC individuals, or used as a vaccine adjuvant. Anti-Zs1g48 antibodies
CC can be produced using antigenic Zs1g48 epitope-bearing peptides
CC such as the present sequence and used to detect Zs1g48 polypeptides.
XX
SQ Sequence 20 AA;

Query Match 19.6%; Score 114; DB 21; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 SATPKHGLECCPPAPPAVT 57
DB 1 SATPKHGLECCPPAPPAVT 20

RESULT 11
AAV79444
ID AAV79444 standard; Peptide: 20 AA.
AC AAV79444;

DT 01-AUG-2000 (first entry)
XX

DE Human secretory protein 48 (Zs1g48) antigenic peptide.

KW Secretory protein 48; Zs1g48; human; leukocyte; proliferation;
KM Immunostimulant; adjuvant; therapy; antigen; immunogen; antibody.
XX

OS Homo sapiens.

PN W0200018796-A2.

PD 06-APR-2000.

PF 01-OCT-1999; 99MO-US22970.

PR 01-OCT-1998; 98US-0164740.

PA (ZYMO) ZYMOGENETICS INC.

PI Lok S, Shepard PO;

DR WPI; 2000-303441/26.

PT Human secretory protein 48 polypeptides and polynucleotides useful for
PT promoting leukocyte proliferation and for treating immunosuppressed
PT individuals -
XX
XX

PS Claim 2; Page 132; 141pp; English.

XX The present sequence is that of an antigenic peptide of human
CC secretory protein 48 (Zs1g48, see also AAV79436-40), a protein that
CC stimulates the proliferation of peripheral blood mononuclear cells,
CC i.e. T-cells, B-cells and monocytes. The invention provides Zs1g48
CC polynucleotides, polypeptides, antibodies and anti-idiotypic
CC antibodies. Zs1g48 stimulates the proliferation of leukocytes in
CC both a mixed leukocyte reaction and in an unmixd leukocyte
CC reaction. It can be used to promote proliferation of peripheral
CC blood leukocytes and to treat low leukocyte counts in individuals
CC (claimed). This can be useful in treating cancer patients whose
CC leukocytes have been depleted by chemotherapy, radiation or illness.
CC Zs1g48 can be administered to patients receiving bone marrow
CC transplants to promote proliferation of leukocytes produced by the
CC immunosuppressed individuals such as the elderly or HIV infected
CC individuals, or used as a vaccine adjuvant. Anti-Zs1g48 antibodies
CC can be produced using antigenic Zs1g48 epitope-bearing peptides
CC such as the present sequence and used to detect Zs1g48 polypeptides.
XX

SQ Sequence 20 AA;

Query Match 17.7%; Score 103; DB 21; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00031;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 TGDGSAKETVSODKRSQCHT 79
DB 1 TGDGSAKETVSODKRSQCHT 20

```

RESULT 12
AB68417
ID   AB68417 standard; Protein; 743 AA.
XX
AC   AB68417;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 32043.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PERE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI: 2001-656860/75.
XX
DR   N-PSDB; ABL12520.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 32043; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid (detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (AB57737-AB872072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   atftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 743 AA;
XX
Query Match 15.3%; Score 89; DB 22; Length 743;
Best Local Similarity 32.4%; Pred. No. 0.66;
Matches 22; Conservative 15; Mismatches 23; Indels 8; Gaps 3;
OY 18 LGLHLPAL-SLPLVTVAGVWSATPKH--GLEQCPAPPPAVTGTGSGAKETVSODK 73
DB 330 MGLGNLPVPTIPMTKVTSPPIRSSRPMQMSQLAQRPPIPSAFLTSSSPSGSLD-----EK 385
OY 74 RSQGHWTWC 81
DB 386 NAAWYNNWC 393
XX
DE Human ORFX ORF1920 polypeptide sequence SEQ ID NO:3840.

```

```

XX
KW   Human; open reading frame; ORFX; detection; cyrostatic; hepatotropic;
KW   vulnery; antiprostatic; antiparkinsonian; immunosuppressant; cardiant;
KW   anticonvulsant; osteopathic; antidiabetic; antidiabetic;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antidiabetic;
KW   antiviral; antibacterial; antifungal; antineumatic; antihypoid;
KW   antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antinflammatory disease; coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999; 99US-0127607.
PR   02-APR-1999; 99US-0127636.
PR   05-APR-1999; 99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shinkets RA, Leach M;
XX
DR   WPI: 2000-602362/57.
XX
DR   N-PSDB; AAC76365.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 11; Page 2996-2997; 5507pp; English.
XX
AC   AACT7446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,
XX   which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX   sequences have activities such as: cyrostatic; hepatotropic; vulnery;
XX   antiprostatic; antiparkinsonian; nootropic; neuroprotective;
XX   osteopathic; anticonvulsant; antidiabetic; immunosuppressant;
XX   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX   antidiabetic; hypotensive; dermatological; immunosuppressive;
XX   antinflammatory; antibacterial; antiviral; antifungal; antineumatic;
XX   antihypoid; and antianemic. The sequences can be used for determining
XX   the presence of or predisposition to, or preventing or treating
XX   pathological conditions associated with an ORFX-associated disorder. The
XX   nucleic acids can be used to express ORFX proteins in gene therapy
XX   vectors. The proteins and nucleic acids may be used to treat cancers,
XX   proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX   graft vs host disease, cardiovascular disease, diabetes mellitus,
XX   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX   nocturnal haemoglobinuria, antinflammatory disease; to enhance
XX   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 252 AA;
XX
Query Match 13.7%; Score 79.5; DB 21; Length 252;
Best Local Similarity 35.1%; Pred. No. 1.9;
Matches 27; Conservative 3; Mismatches 26; Indels 19; Gaps 4;
OY 7 PMPCAHPGLGLGLHLPALSLPLVTVAGVWSAT-----PKHGLEQCPAPPPAVTGTG 61
DB 111 PAVP-----LGGVHPDFQ---VTPGPTGSAADPSWPGHNLHQTGPAPPAVS----- 156

```

OY 62 DSGAKETVSQDKRSQGH 78
 XX :| | | | |
 DB 157 ESMWPLTSGVSPSYSH 173

RESULT 14
 ABB97893

ID ABB97893 standard; Protein: 598 AA.

AC ABB97893;

DT 03-OCT-2002 (first entry)

XX Human secretory polypeptide (SPTM) 145.

XX Human: secretory protein; secretory polynucleotides; SPTM;
 KW SPTM-related disease; somatic gene therapy; germline gene therapy;
 KW severe combined immunodeficiency; intracellular parasite protection;
 KW fungal parasite; protozoan parasite; cell proliferative disorder; cancer;
 KW immune disorder; AIDS; neurological disorder; Parkinson's disease;
 KW motor neuron disorder; demyelinating disease; multiple sclerosis;
 KW meningitis; abscess; prion diseases; cerebral palsy;
 KW neuroskeletal disorder; peripheral nervous system disorder;
 KW dermatomyositis; polymyositis; myopathy; myasthenia gravis;
 KW mental disorder; Tourette's syndrome.

XX Homo sapiens.

PN W0200220756-A2.

PD 14-MAR-2002.

PF 30-AUG-2001; 2001WO-US27297.

XX 05-SEP-2000; 2000US-229747P.
 PR 05-SEP-2000; 2000US-229748P.
 PR 05-SEP-2000; 2000US-229749P.
 PR 05-SEP-2000; 2000US-229750P.
 PR 05-SEP-2000; 2000US-229751P.
 PR 05-SEP-2000; 2000US-230016P.
 PR 05-SEP-2000; 2000US-230583P.
 PR 05-SEP-2000; 2000US-230585P.
 PR 05-SEP-2000; 2000US-230514P.
 PR 05-SEP-2000; 2000US-230515P.
 PR 05-SEP-2000; 2000US-230517P.
 PR 05-SEP-2000; 2000US-230518P.
 PR 05-SEP-2000; 2000US-230519P.
 PR 05-SEP-2000; 2000US-230595P.
 PR 05-SEP-2000; 2000US-230596P.
 PR 05-SEP-2000; 2000US-230597P.
 PR 05-SEP-2000; 2000US-230599P.
 PR 05-SEP-2000; 2000US-230610P.
 PR 05-SEP-2000; 2000US-230864P.
 PR 05-SEP-2000; 2000US-230865P.
 PR 05-SEP-2000; 2000US-230988P.
 PR 05-SEP-2000; 2000US-230989P.
 PR 05-SEP-2000; 2000US-230990P.
 PR 05-SEP-2000; 2000US-230896P.
 PR 07-SEP-2000; 2000US-230897P.
 PR 07-SEP-2000; 2000US-230951P.
 PR 07-SEP-2000; 2000US-231163P.
 PR 07-SEP-2000; 2000US-231833P.

XX (INCYTE) INCYTE GENOMICS INC.

XX Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JT,
 PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR,
 PI Komiya MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM,
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A,
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
 XX MPI; 2002-315658/35.

DR N-PSDB; ABL99890.
 XX Polynucleotide sequences encoding human secretory proteins useful for
 PT gene therapy of e.g. genetic deficiency disorders, cancers, and
 PT diseases caused by intracellular parasites -
 PS Claim 29; Page 459-460; 585pp; English.

CC The invention comprises the amino acid and coding sequences of human
 CC secretory (SPTM) proteins. The SPTM DNA and amino acid sequences are
 CC useful for treating a disease or condition associated with the expression
 CC of functional SPTM. The SPTM DNA sequences are useful for somatic or
 CC germline gene therapy to correct a genetic deficiency (e.g. severe
 CC combined immunodeficiency). The SPTM DNA sequences are also useful in
 CC providing protection against intracellular parasites (e.g. fungal
 CC parasites and protozoan parasites). The SPTM DNA and protein sequences
 CC are also useful for diagnosing cell proliferative disorders, cancer,
 CC immune disorders (e.g. AIDS), neurological disorders (e.g. Parkinson's
 CC disease), motor neuron disorders, demyelinating diseases (e.g. multiple
 CC sclerosis), meningitis, abscesses, prion diseases, cerebral palsy,
 CC neuroskeletal disorders, peripheral nervous system disorders,
 CC dermatomyositis and polymyositis, myopathy, myasthenia gravis, and mental
 CC disorders (e.g. Tourette's syndrome). Amino acid sequences ABB97749 -
 CC ABB97933 represent human secretory proteins of the invention.

XX Sequence 598 AA;

Query Match 13.7%; Score 79.5; DB 23; Length 598;
 Best Local Similarity 29.7%; Pred. No. 5.3;
 Matches 35; Conservative 13; Mismatches 53; Indels 17; Gaps 5;

OY 2 LGYSEPMCAHPGLFL-IGLHPALSLP---LWTVAGVMSATPKHGLQC-----P 49
 DB 34 MGLSAAPLWGPGLLIALHPALSYPRRDYCVLGAGPAGLQMAVFLQRAGRDYAVFE 93

OY 50 PAPPPAVTGTSGSGAKETVSQDKRSQGHWCFLAHPHPLTWGCH---LRHHVSSA 103
 DB 94 RAPPFG-SFETRYPRHKLISINKRYTKGNAEFNLRHDMNSLLSHDPRLLFHHYSNA 150

RESULT 15

ID AAM96103 standard; Protein: 136 AA.

AC AAM96103;

DT 21-NOV-2001 (first entry)

DE Human reproductive system related antigen SEQ ID NO: 4761.

KW Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy.

XX Homo sapiens.

PN W0200155320-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01339.

XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.

	Matches	33;	Conservative	5;	Mismatches	36;	Indels	42;	Gaps	5;
QY	7	PMPCANPLGLFLGLHPALSLPLVVTAGVMSAT-----	PKHGLECCRPAPPAVTGPTG	61						
Db	33	PAPC-----	LGSVHNDFO---VTGPRGTFSADPSPWPGHNLHQTGRAPPAVS-----	78						
QY	62	DSGAKETVSQDKRSQGH-----	TWCTLALPHPWLTWVG	94						
Db	79	ESMPYPLTSQVSPSPSHMDVYMRHPTLMPTCTATATIIITITLLALPWIHPMG		134						

Search completed: June 17, 2003, 10:34:57
 Job time : 71 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2003, 03:24:38 ; Search time 1081 Seconds

(without alignments)
8480.466 Million cell updates/sec

Title: US-09-955-807-1_COPY_59_373

Perfect score: 315
Sequence: 1 atgtctgtatcttgagcc.....atgtctcttcagcagccac 315

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	313.4	99.5	158219	9 AC005534	AC005534 Homo sapi
2	38.8	12.3	136124	9 HS394P21	AL021528 Homo sapi
3	38.8	12.3	155112	2 AL357521	AL357521 Homo sapi
4	38.8	12.3	158297	2 AL513545	AL513545 Homo sapi
5	38.8	12.3	209676	2 AL831755	AL831755 Homo sapi
6	37	11.7	98256	2 AC126501	AC126501 Rattus no
7	36.8	11.7	113666	2 AC105744	AC105744 Oryza sat
8	36.6	11.6	116704	10 AL591921	AL591921 Mouse DNA
9	36.6	11.6	192153	2 AC095187	AC095187 Rattus no
10	36.6	11.5	209882	2 AC103554	AC103554 Rattus no
11	36.2	11.5	3688	10 RMO7628	AJ007628 Rattus no
12	36.2	11.5	3736	6 AR179196	AR179196 Sequence
13	36.2	11.5	3736	10 AB022699	AB022699 Rattus no
14	36.2	11.5	10732	6 E32986	E32986 Gene encodi
15	36.2	11.5	110000	2 AC021632_1	Continuation (2 of
16	36.2	11.5	142802	2 AC127725	AC127725 Rattus no
17	36.2	11.5	145616	9 HS108K11	HS108K11 Human DNA
18	36.2	11.5	152982	2 AC023804	AC023804 Mouse muscu
19	36.2	11.5	168981	10 AL591469	AL591469 Mouse DNA
20	36.2	11.5	263776	2 AC087335	AC087335 Mus muscu
21	35.4	11.2	43012	9 AC079891	AC079891 Homo sapi
22	35.4	11.2	111253	9 HS273P20	AL034371 Human DNA
23	35.4	11.2	185126	2 AC116269	AC116269 Rattus no
24	35.4	11.2	292767	2 AC048333	AC048333 Homo sapi
25	35.2	11.2	300233	2 AC123842	AC123842 Mus muscu
26	35	11.1	164290	2 AC096454	AC096454 Rattus no
27	35	11.1	191815	2 AC097557	AC097557 Rattus no
28	34.8	11.0	99477	9 AL590438	AL590438 Human DNA
29	34.8	11.0	140572	9 AL354743	AL354743 Human DNA
30	34.8	11.0	196329	2 AC129673	AC129673 Rattus no
31	34.6	11.0	154441	2 AC019182	AC019182 Homo sapi
32	34.6	11.0	161443	2 AC011766	AC011766 Homo sapi
33	34.6	11.0	162133	2 AC068569	AC068569 Homo sapi
34	34.6	11.0	174302	2 AC022692	AC022692 Homo sapi
35	34.4	10.9	126095	2 AC104315	AC104315 Rattus no
36	34.4	10.9	128332	9 AC068657	AC068657 Homo sapi
37	34.4	10.9	202559	10 AL645849	AL645849 Mouse DNA
38	34.4	10.9	218904	2 AL844840	AL844840 Mus muscu
39	34.4	10.9	240236	2 AC124765	AC124765 Mus muscu
40	34.2	10.9	51953	2 AC083779	AC083779 Homo sapi
41	34.2	10.9	161355	9 HSDJ60019	AL080314 Human DNA
42	34.2	10.9	184365	2 AC079639	AC079639 Mus muscu
43	34	10.8	109832	2 AC120978	AC120978 Rattus no
44	34	10.8	150691	2 AC131143	AC131143 Rattus no
45	34	10.8	157197	2 AC099042	AC099042 Oryza sat

ALIGNMENTS

RESULT 1
AC005534 158219 bp DNA linear PRI 21-DEC-1999
LOCUS Homo sapiens PAC clone RP5-982E9 from 7q35-q36, complete sequence.
DEFINITION AC005534
ACCESSION AC005534.2 GI:4753272
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 158219)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE 99063792
 PUBMED 9847074
 REFERENCE 2 (bases 1 to 158219)
 AUTHORS Drone, K., Mohlmann, P. and Eldred, J.
 TITLE The sequence of Homo sapiens PAC clone RP5-982E9
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 158219)
 AUTHORS Waterston, R. H.
 TITLE Direct Submission
 JOURNAL Submitted (25-AUG-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 4 (bases 1 to 158219)
 AUTHORS Waterston, R. H.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAY-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 5 (bases 1 to 158219)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (10-JUL-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 6 (bases 1 to 158219)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On May 5, 1999 this sequence version replaced gi:3907500.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_DJ0982E09

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHRG Chromosome 7 Mapping Project (Eric P. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
 http://www.nhgri.nih.gov/DIR/CTB/CHR7, send
 mailto:sgreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
 This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
 The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
 VECTOR: pCVPAC2
 NEIGHBORING SEQUENCE INFORMATION:
 Actual start of this clone is at base position 1 of RP5-982E9;

actual end is at 158219 of RP5-982E9.
 FEATURES
 A transposon was identified in the cloning vector for RP5-982E9.
 Location/Qualifiers
 source 1..158219

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7q35-q36"
 /clone="RP5-982E9"
 /clone_1tb="RPCI-5"
 1..61
 /rpt_family="Alu"
 1033..1134
 /rpt_family="L1"
 1419..1555
 /rpt_family="L1"
 1559..1634
 /rpt_family="L1"
 1647..1950
 /rpt_family="Alu"
 2463..2573
 /rpt_family="L2"
 3108..3253
 /rpt_family="L1"
 3338..3883
 /rpt_family="L1"
 3879..5566
 /rpt_family="L1"
 5567..5869
 /rpt_family="Alu"
 5870..5959
 /rpt_family="L1"
 5947..6693
 /rpt_family="L1"
 6694..7001
 /rpt_family="Alu"
 7002..8787
 /rpt_family="L1"
 8787..11323
 /rpt_family="L1"
 11324..11620
 /rpt_family="Alu"
 11621..13250
 /rpt_family="L1"
 13492..13986
 /rpt_family="L1"
 14003..14189
 /rpt_family="L1"
 14301..15249
 /rpt_family="L1"
 15264..15379
 /rpt_family="L1"
 15410..15690
 /rpt_family="L1"
 15693..15721
 /rpt_family="L1"
 15737..16069
 /rpt_family="MALR"
 16070..16349
 /rpt_family="MALR"
 16353..17739
 /rpt_family="L1"
 17740..18883
 /rpt_family="MALR"
 18884..19913
 /rpt_family="Retroviral"
 19914..20121
 /rpt_family="MALR"
 20122..20459
 /rpt_family="MALR"
 20463..21155
 /rpt_family="L1"

```

repeat_region 21156..21574
                /rpl_family="MallR"
repeat_region 21577..21867
                /rpl_family="Alu"
repeat_region 21868..22465
                /rpl_family="L1"
repeat_region 22466..22732
                /rpl_family="Alu"
repeat_region 22733..23656
                /rpl_family="L1"
repeat_region 23678..24098
                /rpl_family="L1"
repeat_region 24154..24447
                /rpl_family="L1"
repeat_region 24522..24815
                /rpl_family="Alu"
repeat_region 27973..28310
                /rpl_family="L2"
repeat_region 28706..28868
                /note="match to EST AA400700 (NID:52054571) zu70g11.r1"
misc_feature   <28709..32820
                /gene="WUSC:H.DJ0982E09.1"
                /join(<28709..28868,32735..32820)
                /gene="WUSC:H.DJ0982E09.1"
                /note="similar to AAC34385.1 (PID:93170538);
                H.DJ0982E09.1"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAd3188.1"
                /db_xref="GI:5441943"
                /translation="ITRFDLGDGFRFNMIGNFYIVLSNLLFAIVTTLVRRKTS
                VREELFKALGHRKHLPTMSRDSSEFAKPSVNGHOKAL"
repeat_region 29000..29170
                /rpl_family="MIR"
repeat_region 31161..31302
                /rpl_family="MER2_type"
repeat_region 31304..31553
                /rpl_family="MER2_type"
repeat_region 31936..32223
                /rpl_family="Alu"
repeat_region 32262..32573

Query Match      99.5%; Score 313.4; DB 9; Length 158219;
Best Local Similarity 99.7%; Pred. No. 4.2e-88;
Matches 314; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATGCTGGTATTTCTGAGCCCATGGCATGTGCACACCCACCTGGCTTCTCTTAAAGC 60
    |||||||
Db 39257 ATGCTGGTATTTCTGAGCCCATGGCATGTGCACACCCACCTGGCTTCTCTTAAAGC 39316
    |||||||

OY 61 CTACACCTGCGCTTCTTGGCCCTGTACTACTGTGGTGGATGATGAGCCCACT 120
    |||||||
Db 39317 CTACACCTGCGCTTCTTGGCCCTGTACTACTGTGGTGGATGATGAGCCCACT 39376
    |||||||

OY 121 CCCAAGCATGGCTGGAACAATGTCTCTGCCCCCTCCACAGCAGTGAAGATTCACT 180
    |||||||
Db 39377 CCCAAGCATGGCTGGAACAATGTCTCTGCCCCCTCCACAGCAGTGAAGATTCACT 39436
    |||||||

OY 181 GGGAGCTGGGGGCAAGAGACGTGTGCACAGCAAAAGAGAGGAGGCTCACACATGG 240
    |||||||
Db 39437 GGGAGCTGGGGGCAAGAGACGTGTGCACAGCAAAAGAGAGGAGGCTCACACATGG 39496
    |||||||

OY 241 TGTACCTGCGCGCCCTGACCCCATGGCTGACATGGGTGGACACCTCAGAAATATGTG 300
    |||||||
Db 39497 TGTACCTGCGCGCCCTGACCCCATGGCTGACATGGGTGGACACCTCAGAAATATGTG 39556
    |||||||

OY 301 TCTTCAGGAGCCAC 315
    |||||||
Db 39557 TCTTCAGGAGCCAC 39571
    |||||||

RESULT 2
HS394P21/c

```

```

LOCUS      HS394P21      136124 bp      DNA      linear      PRI 23-NOV-1999
DEFINITION Homo sapiens DNA sequence from clone 394P21 on chromosome
            1p36.12-36.13. Contains the PAX7 gene, locus DIS2644, ESTs and
            STSs, complete sequence.
ACCESSION  AL021528
VERSION    AL021528.1 GI:3115987
KEYWORDS   HTG; DIS2644; PAX7.
SOURCE     Homo sapiens.
            Organism      Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 136124)
AUTHORS   Health P.
TITLE     Direct Submission
JOURNAL   Submitted (09-APR-1998) sanger.ac.uk/HGP/Chr1/ Sanger Centre,
            Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On May 7, 1998 this sequence version replaced gi:2815308.
IMPORTANT: This sequence is not the entire insert of clone 394P21.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre chromosome 1
            mapping project. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 394P21 is at 1 in this sequence. The
            true left end of clone 8B22 is at 136017.
            394P21 is from the library RPI3 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.

FEATURES             location/Qualifiers
     source           1..136124
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p36.12-36.13"
                     /clone="Rp3-394P21"
                     /clone_1id="RPI-3"
     repeat_region    271..309
                     /note="13 copies of 3 mer 92 & conserved"
     repeat_region    659..759
                     /note="MIR repeat: matches 155. .54 of consensus"
     repeat_region    2768..2937
                     /note="MIR repeat: matches 262. .86 of consensus"
     repeat_region    2991..3142
                     /note="MIR repeat: matches 202. .49 of consensus"
     repeat_region    3545..3689
                     /note="MIR repeat: matches 45. .196 of consensus"
     repeat_region    6015..6287
                     /note="AluB repeat: matches 9. .287 of consensus"
     repeat_region    6752..7003
                     /note="AluSg repeat: matches 50. .300 of consensus;
                     incomplete repeat"
     repeat_region    7005..7302
                     /note="AluSg repeat: matches 2. .298 of consensus"
     repeat_region    7528..7581
                     /note="18 copies of 3 mer 98 & conserved"
     repeat_region    7583..7874
                     /note="AluX repeat: matches 294. .1 of consensus"
     repeat_region    9738..9998

```

```
/note="AluY repeat: matches 301. .36 of consensus;
incomplete repeat"
misc_feature
/note="match: match GSS B84649"
repeat_region
12521..12704
/note="MIR repeat: matches 207. .18 of consensus"
repeat_region
14315..14478
/note="MIR repeat: matches 48. .217 of consensus"
repeat_region
17372..17523
/note="FAM repeat: matches 16. .170 of consensus"
prim_transcript
17988..18379
/note="match: EST AA582209"
repeat_region
22310..22405
/note="12 copies of 8 mer 91 & conserved"
repeat_region
23505..23709
/note="MIR repeat: matches 51. .256 of consensus"
repeat_region
23807..23934
/note="MIR repeat: matches 70. .192 of consensus"
repeat_region
24357..24654
/note="AluY repeat: matches 3. .301 of consensus"
repeat_region
24705..24917
/note="MIR repeat: matches 256. .40 of consensus"
repeat_region
25341..25643
/note="AluSX repeat: matches 302. .1 of consensus"
repeat_region
26251..26361
/note="MIR repeat: matches 202. .84 of consensus"
repeat_region
27216..27317
/note="MIR repeat: matches 80. .190 of consensus"
repeat_region
27572..27657
/note="MIR repeat: matches 63. .149 of consensus"
repeat_region
27960..28002
/note="MIR repeat: matches 145. .102 of consensus"
repeat_region
28625..28684
/note="3 copies of 20 mer 83 & conserved"
repeat_region
30328..30627
/note="AluSX repeat: matches 300. .1 of consensus"
gene
complement(30672..135804)
/note="PAX7"
mRNA
complement(101n(30672..31178,63514..63716,65992..66157,
74857..75056,130439..130573,131570..131699,132272..132507,
135122..135804))
/note="PAX7"
/product="d1394p21.1 (PAX-7)"
/note="match: CDNAS X66743 Z35141 D87838 AF014369; match:
genomic DNAs X96748 X96747 X96746 X96745 X15042 X15250
X15251; polyA tail at 30672"
/evidence="not_experimental"
complement(101n(30771..31178,63514..63716,65992..66157,
74857..75056,130439..130573,131570..131699,132272..132507,
135122..135206))
/note="PAX7"
/note="match: proteins P23759 Q42349 P47239"
/codon_start=-1
/evidence="not_experimental"
/product="d1394p21.1 (PAX-7)"
/protein_id="CAA16432.1"
/db_xref="GI:3115968"
/translation="MAALPGTVPRMRRPAPGQNYPTGPLEVSTPLGQVGNOLGV
FINGRPLPNHRIKIVEMAHGIRPCVISLIVSGCVSKICRQETSIRPGAIG
GSGRQVAPDVPEKITEEKKRNGPMSWIRDLKDKGCHDSILCRQVSSISRL
RIKFGKKEEDEADKDEKKAHSIOLIGDKNRKIDEGSDVSEPLPKROR
RSRTTTAOLDELEKAFERTHPDITRELLAKRKLREARQVFNRRARRKA
GANQLAENHLLPGGFPPTGMPPLPYQLPDSTYPTTTISQDGSVHRPQLPSTM
HOGGLAANAADATSSAYGARHSFSSYSOSFMPAPSNMHNNSGLSPQVMSILGN
PSAVPPQADAFESISPLHGIDSATSSISQADSIKRGDSLPTSOACQPTYSST
GKSVDPVAGQYQGSQSLVPMASVPPIPSPTPRASCLFMSYVSVSWMSISQM
EKLSQMEQET"
repeat_region
31490..31529
/note="10 copies of 4 mer 90 & conserved"
repeat_region
32310..32419
/note="2 copies of 55 mer 96 & conserved"
repeat_region
32827..32951
/note="MIR repeat: matches 196. .80 of consensus"
```

```
repeat_region
34330..34448
/note="MIR2 repeat: matches 24. .143 of consensus"
repeat_region
34452..34589
/note="MIR2 repeat: matches 142. .1 of consensus"
repeat_region
35288..35442
/note="MIR repeat: matches 154. .2 of consensus"
repeat_region
36533..36718
/note="MIR repeat: matches 66. .258 of consensus"
repeat_region
37399..37453
/note="MIR repeat: matches 139. .89 of consensus"
repeat_region
37927..38231
/note="AluSP repeat: matches 301. .1 of consensus"
repeat_region
38385..38486
/note="2 copies of 51 mer 96 & conserved"
repeat_region
39141..39290
/note="MIR repeat: matches 211. .50 of consensus"
repeat_region
39398..39475
/note="3 copies of 26 mer 81 & conserved"
repeat_region
39598..39625
/note="Single clone region"
repeat_region
40254..40316
/note="3 copies of 21 mer 91 & conserved"
repeat_region
40337..40387
/note="MIR2 repeat: matches 95. .146 of consensus"
repeat_region
40626..40669
/note="22 copies of 2 mer 93 & conserved"
repeat_region
40760..40829
/note="35 copies of 2 mer 81 & conserved"
repeat_region
41713..41738
/note="MIR repeat: matches 117. .143 of consensus"
repeat_region
42043..42137
/note="MIR repeat: matches 143. .49 of consensus"
repeat_region
44660..44845
/note="MIR repeat: matches 256. .74 of consensus"
repeat_region
46142..46240
/note="3 copies of 33 mer 93 & conserved"
repeat_region
47871..48166
/note="AluSG repeat: matches 1. .297 of consensus"
repeat_region
48341..48435
/note="MERSA repeat: matches 55. .148 of consensus"
repeat_region
49084..49131
/note="MIR2 repeat: matches 132. .85 of consensus"
repeat_region
49749..49775
/note="9 copies of 3 mer 93 & conserved"
repeat_region
49782..49837
/note="51237
51237
51254..51431
/note="MIR repeat: matches 230. .2 of consensus"
repeat_region
51568..52245
/note="MIR repeat: matches 88. .262 of consensus"
repeat_region
52370..52562
/note="MIR repeat: matches 142. .63 of consensus"
repeat_region
52585..52880
/note="MIR repeat: matches 252. .55 of consensus"
```

```
Query Match 12.38; Score 38.8; DB 9; Length 136124;
Best Local Similarity 57.48; Pred. NO. 0.34;
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 77 CTTGGCCCTTGAGTACTGGCGGAGTGAATGAGCCGCACTCCCAACAGGCGCTGG 136
DB 92039 CGTGGTCTCTTTATTTGCGANGCTGGTTTCTCTGCATTTCTATAGAACAGTGTTCT 91980
QY 137 ACAATGTCCTCCCTGCCCTCCACAGCAGTGAAGATTCAGTGGGAGTCGGGGCAA 196
DB 91979 CAATATGTGACCAACCCCAACAGCAGATGCGACGATCATCTGGAACTTCTGGAA 91920
QY 197 AG 198
DB 91919 TG 91918
RESULT 3
```

```

AL357521      155112 bp      DNA      linear      HTG 10-JUL-2001
LOCUS
DEFINITION   Homo sapiens chromosome 1 clone RP5-926022, *** SEQUENCING IN
PROGRESS ***; 20 unordered pieces.
ACCESSION
AL357521
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155112)
McJarry, K.
Direct Submission
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonequests@sanger.ac.uk
On Nov 23, 2000 this sequence version replaced gi:10039960.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: dr926022
----- Summary Statistics
Assembly program: XGAP; version 4.5
Sequencing vector: M13; M77815; 7% of reads
Sequencing vector: plasmid; L08752; 92% of reads
Chemistry: Dye-terminator ET-amersham; 12% of reads
Chemistry: Dye-terminator Big Dye; 68% of reads
Chemistry: Dye-terminator ABI; 19% of reads
Consensus quality: 145025 bases at least Q40
Consensus quality: 148804 bases at least Q30
Consensus quality: 151286 bases at least Q20
Insert size: 153212; sum-of-contigs
Insert size: 121507; 87.0% error; agarose-1p
Quality coverage: 5.16x in Q20 bases; sum-of-contigs quality
coverage: 7.34x in Q20 bases; agarose-1p
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
8489 8588: contig of 8488 bp in length
8489 8588: gap of 100 bp
8589 14083: contig of 5495 bp in length
14084 14183: gap of 100 bp
14184 16498: contig of 2315 bp in length
16499 16598: gap of 100 bp
16599 223796: contig of 7200 bp in length
23799 23898: gap of 100 bp
23899 36662: contig of 12764 bp in length
36663 36762: gap of 100 bp
36763 45878: contig of 9116 bp in length
45879 45978: gap of 100 bp
45979 50067: contig of 4089 bp in length
50068 50167: gap of 100 bp
50168 56588: contig of 6421 bp in length
56589 56688: gap of 100 bp
56689 63582: contig of 6894 bp in length
63583 63682: gap of 100 bp
63683 68434: contig of 4752 bp in length
68435 68534: gap of 100 bp
68535 75603: contig of 7069 bp in length
75604 75703: gap of 100 bp
75704 78061: contig of 2358 bp in length
78062 78161: gap of 100 bp
78162 88954: contig of 10793 bp in length

```

	*	88955	89054:	gap of	100	bp
	*	89055	92586:	contig of	3532	bp in length
	*	92587	92686:	gap of	100	bp
	*	92687	96459:	contig of	3773	bp in length
	*	96460	96559:	gap of	100	bp
	*	96560	101475:	contig of	4916	bp in length
	*	101476	101575:	gap of	100	bp
	*	101576	115520:	contig of	13945	bp in length
	*	115521	115620:	gap of	100	bp
	*	115621	146532:	contig of	30912	bp in length
	*	146533	146632:	gap of	100	bp
	*	146633	151919:	contig of	5287	bp in length
	*	151920	152019:	gap of	100	bp
	*	152020	155112:	contig of	3093	bp in length.
FEATURES				Location/Qualifiers		
source				1..155112		
				/organism="Homo sapiens"		
				/db_xref="taxon:9606"		
				/chromosome="1"		
				/clone="RP5-926022"		
				/clone_lib="RPCL-5"		
				1..8488		
misc_feature				/note="assembly_fragment:00871		
				clone_end:sp6		
				vector_side:left"		
misc_feature				8589..14083		
				/note="assembly_fragment:01703		
				fragment_chain:1"		
misc_feature				14184..16498		
				/note="assembly_fragment:02705		
				fragment_chain:1"		
misc_feature				16599..23798		
				/note="assembly_fragment:00068		
				fragment_chain:1"		
misc_feature				23899..36662		
				/note="assembly_fragment:02777		
				fragment_chain:1"		
misc_feature				36763..45878		
				/note="assembly_fragment:00748		
				fragment_chain:2"		
misc_feature				45979..50067		
				/note="assembly_fragment:00024		
				fragment_chain:2"		
misc_feature				50168..56588		
				/note="assembly_fragment:02548		
				fragment_chain:2"		
misc_feature				56689..63582		
				/note="assembly_fragment:01501		
				fragment_chain:3"		
misc_feature				63683..68434		
				/note="assembly_fragment:01410		
				fragment_chain:3"		
misc_feature				68535..75603		
				/note="assembly_fragment:00219"		
misc_feature				75704..78061		
				/note="assembly_fragment:00724"		
misc_feature				78162..88954		
				/note="assembly_fragment:01966"		
misc_feature				89055..92586		
				/note="assembly_fragment:01996"		
misc_feature				92687..96439		
				/note="assembly_fragment:02016"		
misc_feature				96560..101475		
				/note="assembly_fragment:02131"		
misc_feature				101576..115520		
				/note="assembly_fragment:02503"		
misc_feature				115621..146532		
				/note="assembly_fragment:02059		
				fragment_chain:4		
				clone_end:t7		
				vector_side:right"		
misc_feature				146633..151919		
				/note="assembly_fragment:00729		

```

fragment_chain:4"
152020..151112
/misc-feature /note="assembly_fragment:01146
fragment_chain:4
BASE COUNT 38677 a 38785 c 38240 g 37496 t 1914 others
ORIGIN
Query Match 12.3% Score 38.8; DB 2; Length 151112;
Best Local Similarity 57.4%; Prod. No. 0.34; 52; Indels 0; Gaps 0;
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
OY 77 CTTTGCCCTTGATGATTACTGTGCGCTGAGCTGATGAGCCGACCTCCCAACGATGGCTGG 136
Db 88691 CGTGGCTCTGTTATTATGCTGATGCTGGTTTCTCTGCAATTTCTATAGAACATGGTTCT 887500
OY 137 AACAAATGCTCTCGCCCTTCACACGACATGACAGATTTCATGTGGGAGCTGGGGGCA 196
Db 88751 CAATTGTGAAACACCCCAACACACAGATGACATCATCATCTGGAACCTGCTGGAA 888100
OY 197 AG 198
Db 88811 TG 88812
RESULT 4
AL513545 158297 bp DNA linear HMG 10-JUL-2001
LOCUS Homo sapiens chromosome 1 clone RP11-514C23, *** SEQUENCING IN
DEFINITION PROGRESS ***, 18 unordered pieces.
ACCESSION AL513545
VERSION AL513545.2 GI:13160358
KEYWORDS HMG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 158297)
McJAY K.
Direct Submission
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Feb 26, 2001 this sequence version replaced gi:12750921.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA514C23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 153889 bases at least Q40
Consensus quality: 154977 bases at least Q30
Consensus quality: 155599 bases at least Q20
Insert size: 165597; sum-of-contigs
Insert size: 165270; 3.0% error; agarose-fp
Quality coverage: 6.65x in Q20 bases; sum-of-contigs quality
coverage: 6.38x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 4317 4416: gap of 100 bp
* 4417 13709: contig of 9293 bp in length

```

	*	13710	13809:	gap of	100	bp	
	*	13810	16011:	contig of	2202	bp	in length
	*	16012	16111:	gap of	100	bp	
	*	16121	19225:	contig of	3114	bp	in length
	*	19226	19325:	gap of	100	bp	
	*	19326	24319:	contig of	4994	bp	in length
	*	24320	24419:	gap of	100	bp	
	*	24420	32439:	contig of	8020	bp	in length
	*	32540	32539:	gap of	100	bp	
	*	35377	35476:	gap of	100	bp	
	*	35477	48033:	contig of	12557	bp	in length
	*	48034	48133:	gap of	100	bp	
	*	48134	65466:	contig of	17333	bp	in length
	*	65467	65566:	gap of	100	bp	
	*	65567	70347:	contig of	4781	bp	in length
	*	70348	70447:	gap of	100	bp	
	*	70448	87906:	contig of	17459	bp	in length
	*	87907	88006:	gap of	100	bp	
	*	88007	96207:	contig of	8201	bp	in length
	*	96308	96307:	gap of	100	bp	
	*	96308	98782:	contig of	2475	bp	in length
	*	98783	98882:	gap of	100	bp	
	*	98883	101526:	contig of	2644	bp	in length
	*	101527	101626:	gap of	100	bp	
	*	101627	130171:	contig of	28545	bp	in length
	*	130172	132071:	gap of	100	bp	
	*	130272	140166:	contig of	9855	bp	in length
	*	140167	140266:	gap of	100	bp	
	*	140267	152701:	contig of	12435	bp	in length
	*	152702	152801:	gap of	100	bp	
	*	152802	158297:	contig of	5456	bp	in length.
FEATURES							
source							
			1.	158297			
			/organism="Homo sapiens"				
			/db_xref="taxon:9606"				
			/chromosome="1"				
			/clone="RP11-514C23"				
			/clone_11b="RPc1-11.2"				
			1.	4316			
misc_feature			/note="assembly-fragment:02586				
			fragment_chain:1				
			clone_end:SP6				
			vector_side:left"				
misc_feature			4417.	13709			
			/note="assembly-fragment:01913				
			fragment_chain:1"				
misc_feature			13810.	16011			
			/note="assembly-fragment:01793				
			fragment_chain:1"				
misc_feature			16112.	19225			
			/note="assembly-fragment:02161				
			fragment_chain:1"				
misc_feature			19326.	24319			
			/note="assembly-fragment:00039				
			fragment_chain:1"				
misc_feature			24420.	32439			
			/note="assembly-fragment:01183				
			fragment_chain:1"				
misc_feature			32540.	35376			
			/note="assembly-fragment:01645				
			fragment_chain:1"				
misc_feature			35477.	48033			
			/note="assembly-fragment:01800				
			fragment_chain:2"				
misc_feature			48134.	65466			
			/note="assembly-fragment:02272				
			fragment_chain:2"				
misc_feature			65567.	70347			
			/note="assembly-fragment:02380				
			fragment_chain:2"				
misc_feature			70448.	87906			
			/note="assembly-fragment:00477				


```

misc_feature      fragment_chain:2"
88007..96207
/Note="assembly-fragment:00137
fragment_chain:3"
misc_feature      /Note="assembly-fragment:01506
96308..98782
/Note="assembly-fragment:01506
fragment_chain:3"
misc_feature      /Note="assembly-fragment:00660
98883..101526
/Note="assembly-fragment:00660
fragment_chain:3"
misc_feature      /Note="assembly-fragment:00061
101627..130171
/Note="assembly-fragment:00061
fragment_chain:3"
misc_feature      /Note="assembly-fragment:01928
130272..140166
/Note="assembly-fragment:01928
fragment_chain:3"
misc_feature      /Note="assembly-fragment:00148
140267..152701
/Note="assembly-fragment:00148
fragment_chain:3"
misc_feature      /Note="assembly-fragment:0152802
152802..158297
/Note="assembly-fragment:0152802
fragment_chain:3"
BASE COUNT      39009 a 38793 c 40866 g 37912 t 1717 others
ORIGIN

```

```

Query Match      12.3%; Score 38.8; DB 2; Length 158297;
Best Local Similarity 57.4%; Pred. No. 0.34;
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 77 CTTGGCCCTTGTAGTACTGTGCTGAGTATGATGAGCCGACCTCCAGCAGTGGCTCG 136
DB 46294 CTTGGCCCTTGTATTTGCTGATGCTGCTTTCTCTGCACTTCTATAGACAGTGTCT 46353
QY 137 AACATGTCCTCTGCTGCTCCACGACGAGTACAGATTCATCTGGGAGCTGGGGGCAA 196
DB 46354 CAAATGTGACACCCGCCACGACGAGTGCACATCTGGGAATCTGCTGGGAA 46413
QY 197 AG 198
DB 46414 TG 46415

```

```

RESULT 5
AL831755      209676 bp      DNA      linear      HTG 09-AUG-2002
LOCUS      Homo sapiens chromosome 1 clone RP13-279N23, *** SEQUENCING IN
DEFINITION
ACCESSION      AL831755
VERSION      AL831755.6 GI:22204771
KEYWORDS      HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 209676)
REFERENCE
AUTHORS      Harrison, E.
TITLE      Direct Submission
JOURNAL      Submitted (02-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonequests@sanger.ac.uk
On Aug 11, 2002 this sequence version replaced gi:21998292.
COMMENT
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: hb279N23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads

```

```

Chemistry: Dye-terminator; 97% of reads
Chemistry: Dye-terminator Big Dye; 2% of reads
Consensus quality: 209220 bases at least Q40
Consensus quality: 209374 bases at least Q30
Consensus quality: 209478 bases at least Q20
Insert size: 209676; sum-of-contigs
Insert size: 218959; 3.2% error; agarose-fp
Quality coverage: 12.24x in Q20 bases; sum-of-contigs Quality
coverage: 11.72x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
source
1..209676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP13-279N23"
/clone_1lb="RPCT-13.2"
1..209676
/Note="assembly-fragment:03689
clone_end:SP6
vector_side:right"

```

```

BASE COUNT      52994 a 53379 c 52820 g 50483 t
ORIGIN

```

```

Query Match      12.3%; Score 38.8; DB 2; Length 209676;
Best Local Similarity 57.4%; Pred. No. 0.34;
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 77 CTTGGCCCTTGTAGTACTGTGCTGAGTATGATGAGCCGACCTCCAGCAGTGGCTCG 136
DB 28788 CTTGGCCCTTGTATTTGCTGATGCTGCTTTCTCTGCACTTCTATAGACAGTGTCT 28847
QY 137 AACATGTCCTCTGCTGCTCCACGACGAGTACAGATTCATCTGGGAGCTGGGGGCAA 196
DB 28848 CAAATGTGACACCCGCCACGACGAGTGCACATCTGGGAATCTGCTGGGAA 28907
QY 197 AG 198
DB 28908 TG 28909

```

```

RESULT 6
AC126501      98256 bp      DNA      linear      HTG 24-JUL-2002
LOCUS      Rattus norvegicus clone CH230-206K3, *** SEQUENCING IN PROGRESS
DEFINITION
ACCESSION      AC126501
VERSION      AC126501.2 GI:21703477
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 98256)
REFERENCE
AUTHORS      Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alshrocks, S.L., Amaral, H.C., Are, J.R., Ayala, M., Banks, T.,
Barbieri, J., Benton, J., Bimge, K., Blankenburg, R., Bonin, D.,
Bouch, J., Bowler, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Bunay, C., Burch, P., Burkett, C., Butrell, K.L., Byrd, N.C.,
Carton, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dethorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Doutlatte, K.U., Draper, H., Dugan-Rocha, S., Durbin, K.D.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

```

Correll, J. H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, D.,
Hennrich, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L. E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisse, H.,
Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metker, M.,
Miner, G., Miner, Z., Mitchell, T., Monabhat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokwen, S., Oguh, M., Okunolu, G.,
Ogunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rotubokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherrer, S., Scott, G., Shen, H., Shoshitani, N., Sisson, I.,
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Swalek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Mieczys, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, D., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 98256)
Worley, K. C.

Direct Submission
Submitted (06-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 98256)
Worley, K. C.

Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 8, 2002 this sequence version replaced gl.21700456.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: G2EG
Center clone name: CH230-206K3
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 67632 bases at least Q40
Consensus quality: 71592 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbankdraft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1161: contig of 1161 bp in length
* 1162 1261: gap of unknown length
* 1262 2480: contig of 1219 bp in length
* 2481 2580: gap of unknown length
* 2581 4115: contig of 1536 bp in length
* 4117 4216: gap of unknown length

4217 5250: contig of 1034 bp in length
* 5251 5350: gap of unknown length
* 5351 5942: contig of 1592 bp in length
* 5943 7042: gap of unknown length
* 7043 8151: contig of 1109 bp in length
* 8152 8251: gap of unknown length
* 8252 10134: contig of 1883 bp in length
* 10135 10234: gap of unknown length
* 10235 11591: contig of 1357 bp in length
* 11592 11692: gap of unknown length
* 11693 13323: contig of 1532 bp in length
* 13324 13423: gap of unknown length
* 13424 15067: contig of 1644 bp in length
* 15068 15167: gap of unknown length
* 15168 17854: contig of 2687 bp in length
* 17855 17954: gap of unknown length
* 17955 20159: contig of 2205 bp in length
* 20160 20259: gap of unknown length
* 20260 23535: contig of 3376 bp in length
* 23536 23635: gap of unknown length
* 23636 26039: contig of 2404 bp in length
* 26040 26139: gap of unknown length
* 26140 28253: contig of 2114 bp in length
* 28254 28353: gap of unknown length
* 28354 31410: contig of 3057 bp in length
* 31411 31510: gap of unknown length
* 31511 33233: contig of 1723 bp in length
* 33234 33333: gap of unknown length
* 33334 36072: contig of 2739 bp in length
* 36073 36172: gap of unknown length
* 36173 38831: contig of 3659 bp in length
* 38832 39931: gap of unknown length
* 39932 44359: contig of 4428 bp in length
* 44360 44459: gap of unknown length
* 44460 49643: contig of 5184 bp in length
* 49644 49743: gap of unknown length
* 49744 53469: contig of 3726 bp in length
* 53470 53569: gap of unknown length
* 53570 57539: contig of 3970 bp in length
* 57540 57639: gap of unknown length
* 57640 64329: contig of 6690 bp in length
* 64330 64429: gap of unknown length
* 64430 68879: contig of 5450 bp in length
* 68880 69979: gap of unknown length
* 69980 74410: contig of 4431 bp in length
* 74411 74510: gap of unknown length
* 74511 81214: contig of 6704 bp in length
* 81215 81314: gap of unknown length
* 81315 87132: contig of 3818 bp in length
* 87133 87232: gap of unknown length
* 87233 91668: contig of 4436 bp in length
* 91669 91768: gap of unknown length
* 91769 98256: contig of 6488 bp in length.

FEATURES
SOURCE
1. 98256
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-206K3"

BASE COUNT 24160 a 23770 c 23993 g 23334 t 2999 others

ORIGIN
Query Match 11.7% Score 37; DB 2; Length 98256;
Best Local Similarity 53.0%; Pred. No. 1.3;
Matches 79; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 13 TCTGAGCCATGCATGTCACACCCACTTGGCTTCTTCTTAGGCTTACACCTTGGC 72
DB 22156 TCTGAATTCACATCTTCTGATCTATTCCGACACAGAGCATATAGGCGATGATGC 132
QY 73 CTTTCTTGGCCCTTGTAGTACGTGCGTGAAGTGAAGCGCCTCCCAAGCATGGC 132
DB 22216 GTCTCTTCCCTTGTCTTCTTGTCTGGGCTGACAGGATAAACGAGACCCCAAGCATCTC 22275

[illegible]

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 192153)
Morley, K. C.
Submitted (11-Jul-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi.17942286.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GDBX
Center clone name: CH230-9C3

----- Summary Statistics -----
Sequencing Vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 119125 bases at least Q40
Consensus quality: 125091 bases at least Q30
Consensus quality: 130301 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drafc_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 79 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	1243:	contig of 1243 bp in length
*	1244	1343:	gap of unknown length
*	1344	2721:	contig of 1378 bp in length
*	2722	2821:	gap of unknown length
*	2822	3941:	contig of 1120 bp in length
*	3942	4041:	gap of unknown length
*	4042	5289:	contig of 1248 bp in length
*	5290	5389:	gap of unknown length
*	5390	6956:	contig of 1567 bp in length
*	6957	7056:	gap of unknown length
*	7057	8133:	contig of 1077 bp in length
*	8134	8233:	gap of unknown length
*	8234	9263:	contig of 1032 bp in length
*	9266	9365:	gap of unknown length
*	9366	10944:	contig of 1579 bp in length
*	10945	11044:	gap of unknown length
*	11045	12750:	contig of 1706 bp in length
*	12751	12850:	gap of unknown length
*	12851	13958:	contig of 1108 bp in length
*	13959	14058:	gap of unknown length
*	14059	15275:	contig of 1217 bp in length
*	15276	15375:	gap of unknown length
*	15376	16793:	contig of 1418 bp in length
*	16794	16893:	gap of unknown length
*	16894	18019:	contig of 1126 bp in length
*	18020	18119:	gap of unknown length
*	18120	19768:	contig of 1649 bp in length
*	19769	19868:	gap of unknown length
*	19869	21147:	contig of 1279 bp in length
*	21148	21247:	gap of unknown length
*	21248	23000:	contig of 1753 bp in length
*	23001	23100:	gap of unknown length
*	23101	25073:	contig of 1973 bp in length
*	25074	25173:	gap of unknown length
*	25174	27094:	contig of 1921 bp in length
*	27095	27194:	gap of unknown length
*	27195	28757:	contig of 1563 bp in length
*	28758	28857:	gap of unknown length
*	28858	30242:	contig of 1385 bp in length

```

* 30243 30342: gap of unknown length
* 30343 32113: contig of 1771 bp in length
* 32114 32213: gap of unknown length
* 32214 34175: contig of 1962 bp in length
* 34176 34275: gap of unknown length
* 34276 35470: contig of 1195 bp in length
* 35471 37322: gap of unknown length
* 37323 37422: gap of unknown length
* 37423 38791: contig of 1369 bp in length
* 38792 38891: gap of unknown length
* 38892 40361: contig of 1470 bp in length
* 40362 40461: gap of unknown length
* 40462 41995: contig of 1534 bp in length
* 41996 42095: gap of unknown length
* 42096 44157: contig of 2062 bp in length
* 44158 44257: gap of unknown length
* 44258 45517: contig of 1260 bp in length
* 45518 45617: gap of unknown length
* 45618 46971: contig of 1353 bp in length
* 46971 47070: gap of unknown length
* 47071 49593: contig of 2523 bp in length
* 49594 49693: gap of unknown length
* 49694 51492: contig of 1799 bp in length
* 51493 51592: gap of unknown length
* 51593 53277: contig of 1685 bp in length
* 53278 53377: gap of unknown length
* 53378 54782: contig of 1405 bp in length
* 54783 54882: gap of unknown length
* 54883 57021: contig of 2139 bp in length
* 57022 57121: gap of unknown length
* 57122 59001: contig of 1880 bp in length
* 59002 59101: gap of unknown length
* 59102 60121: contig of 1020 bp in length
* 60122 60221: gap of unknown length
* 60222 62559: contig of 2338 bp in length
* 62560 62659: gap of unknown length
* 62660 64766: contig of 2107 bp in length
* 64767 64866: gap of unknown length
* 64867 66886: contig of 1820 bp in length
* 66887 66786: gap of unknown length
* 66787 68406: contig of 1620 bp in length
* 68407 68506: gap of unknown length
* 68507 70230: contig of 1724 bp in length
* 70231 70330: gap of unknown length
* 70331 72687: contig of 2357 bp in length
* 72688 72787: gap of unknown length
* 72788 73961: contig of 1174 bp in length
* 73962 74061: gap of unknown length
* 74062 75594: contig of 1533 bp in length
* 75595 75694: gap of unknown length
* 75695 77376: contig of 1682 bp in length
* 77377 77476: gap of unknown length
* 77477 79269: contig of 1793 bp in length
* 79270 79370: gap of unknown length
* 79371 81510: contig of 2141 bp in length
* 81511 81610: gap of unknown length
* 81611 83780: contig of 2170 bp in length
* 83781 83880: gap of unknown length
* 83881 86085: contig of 2205 bp in length
* 86086 86185: gap of unknown length
* 86186 88050: contig of 1865 bp in length
* 88051 88150: gap of unknown length
* 88151 91051: contig of 2901 bp in length
* 91052 91151: gap of unknown length
* 91152 93335: contig of 2184 bp in length
* 93336 93435: gap of unknown length
* 93436 96781: contig of 3346 bp in length

```

Query Match 11.6%; Score 36.6; DB 2; Length 192153;
 Best Local Similarity 50.9%; Pred. No. 1.7;
 Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```

QY 10 TATTGTGAGCCCATGATGACACCCACTGTGGCTTCCTCTTGTAGGCTACACCT 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40962 TATGAGGTGCCCCCCCCACACACACCCCTCCACCCCTCCCATCCACCAT 40903
QY 70 GCCCTTTTGGCCCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40902 GACATGCTTCACTATGAGGATGATGATGATGATGATGATGATGATGATGAT 40843
QY 130 GCCCTGGAACATGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40842 GTCCAAACAAGACCGCTCTACTACATATGCAACTGAAGTATGGTCCCT 40792

RESULT 10
LOCUS AC103554
DEFINITION Rattus norvegicus clone CH230-120K20, *** SEQUENCING IN PROGRESS
ACCESSION AC103554.6 GI:22094267
VERSION AC103554.6
KEYWORDS HTG, HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 209882)
AUTHORS Muzny,D,Marie, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alsbrooks,S, Amin,A, Angiano,D,
Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Cesar,H, Center,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,D,
Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Davila,M,L, Davis,C, Davy-Carroll,L, De Andrade,C, Dederich,D,
Delgado,O, Denison,S, Derramo,C, Ding,Y, Dinh,H, Divya,K,
Drepper,H, Dugan-Rocha,S, Dunn,A, Durkin,K, Duval,B, Evans,K,
Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
Georgiadis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W,
Gunatirane,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K,
Harvey,Y, Havlak,P, Hayes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M,
Hollins,B, Howells,S, Huiyk,S, Hume,J, Idlebird,D, Jackson,A,
Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorensunewa,L, Louiseged,H, Lozada,R,U, Lu,X, Ma,J,
Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A,
Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E,
Mawney,S, McLeod,M, McNeill,J, Meenen,E, Milosavljevic,A,
Miner,G, Minja,E, Montemayor,T, Moore,S, Morgan,M, Morris,K,
Morris,S, Munidasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D,
Newton,N, Nguyen,N, Norris,S, Nwokediemen,O, Okunnu,G,
Olanunpunsagoun,A, Pal,S, Parks,K, Pasternak,S, Paul,H,
Perez,A, Perez,L, Pfannkuch,C, Plopper,F, Polidexter,A,
Popovic,D, Primus,E, Pu,L, Puazo,M, Quito,J, Rachlin,E,
Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y,
Reuter,M, Richards,S, Riggs,P, Rivers,C, Rocket,T, Rojas,A,
Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savery,G, Scherer,S,
Scott,G, Shatsman,S, Shen,H, Shetty,U, Shvartsbeyn,A,
Sisong,I, Silter,C,D, Smaj,D, Sneed,A, Sodergren,E,
Song,X,-Z, Sorrelle,R, Sosa,D, Steimle,M, Strong,R, Sutton,A,
Svatek,A, Tabbor,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S,
Tingey,A, Trejos,Z, Usmani,K, Valas,R, Vera,Y, Villasana,D,
Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,R,
Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wleczek,R,
Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S,
Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X,

```

TITLE Zhao S., Dunn D., von Niederhausern A., Weiss R., Smith D.R.,
 JOURNAL Holt R.A., Smith H.O., Weinstock G. and Gibbs R.A.
 REFERENCE Unpublished
 AUTHORS 2 (bases 1 to 209882)
 TITLE Morley, K.C.
 JOURNAL Direct Submission
 REFERENCE Submitted (28-NOV-2001) Human Genome Sequencing Center; Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 209882)
 Rat Genome Sequencing Consortium.
 Direct Submission
 REFERENCE Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Aug 3, 2002 this sequence version replaced gi:22091298.
 COMMENT ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GGFP
 Center clone name: CH230-120K20
 Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 167910 bases at least Q40
 Consensus quality: 174235 bases at least Q30
 Consensus quality: 178487 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 60 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 1129: contig of 1129 bp in length
 1130 1229: gap of unknown length
 1230 3065: contig of 1836 bp in length
 3066 3165: gap of unknown length
 3166 4193: contig of 1027 bp in length
 4193 4293: gap of unknown length
 4293 5851: gap of 1559 bp in length
 5852 5951: gap of unknown length
 5952 7127: contig of 1176 bp in length
 7128 7227: gap of unknown length
 7228 8256: contig of 1029 bp in length
 8257 8357: gap of unknown length
 8357 9467: contig of 1111 bp in length
 9468 9567: gap of unknown length
 9568 10838: contig of 1271 bp in length
 10839 10938: gap of unknown length
 10939 12068: contig of 1130 bp in length
 12069 12168: gap of unknown length
 12169 13197: contig of 1028 bp in length
 13197 13296: gap of unknown length
 13297 14315: contig of 1019 bp in length
 14316 14415: gap of unknown length
 14416 15738: contig of 1323 bp in length
 15739 15838: gap of unknown length
 15839 17887: contig of 2049 bp in length
 17888 17987: gap of unknown length
 17988 19728: contig of 1741 bp in length
 19729 19828: gap of unknown length
 19829 21303: contig of 1475 bp in length
 21304 21403: gap of unknown length

*	21404	22678:	contig of 1275 bp in length
*	22679	22778:	gap of unknown length
*	22779	23955:	contig of 1177 bp in length
*	23956	24055:	gap of unknown length
*	24056	26296:	contig of 2241 bp in length
*	26297	26396:	gap of unknown length
*	26397	27535:	contig of 1139 bp in length
*	27536	27635:	gap of unknown length
*	27636	29582:	contig of 1947 bp in length
*	29583	29683:	gap of unknown length
*	29683	31852:	contig of 2170 bp in length
*	31853	31952:	gap of unknown length
*	31953	33524:	contig of 1572 bp in length
*	33525	33624:	gap of unknown length
*	33625	36407:	contig of 2783 bp in length
*	36408	36507:	gap of unknown length
*	36508	38752:	contig of 2245 bp in length
*	38753	38852:	gap of unknown length
*	38853	40898:	contig of 2046 bp in length
*	40899	40998:	gap of unknown length
*	40999	43289:	contig of 2291 bp in length
*	43290	43389:	gap of unknown length
*	43390	45173:	contig of 1784 bp in length
*	45174	45273:	gap of unknown length
*	45274	47552:	contig of 2279 bp in length
*	47553	47652:	gap of unknown length
*	47653	49457:	contig of 1805 bp in length
*	49458	49557:	gap of unknown length
*	49558	52516:	contig of 2959 bp in length
*	52517	52617:	gap of unknown length
*	52617	55579:	contig of 2963 bp in length
*	55580	55679:	gap of unknown length
*	55680	57710:	contig of 2031 bp in length
*	57711	57810:	gap of unknown length
*	57811	60388:	contig of 2578 bp in length
*	60389	60488:	gap of unknown length
*	60489	63297:	contig of 2809 bp in length
*	63298	63397:	gap of unknown length
*	63398	67352:	contig of 3955 bp in length
*	67353	67452:	gap of unknown length
*	67453	71221:	contig of 3769 bp in length
*	71222	71321:	gap of unknown length
*	71322	74908:	contig of 3587 bp in length
*	74909	75008:	gap of unknown length
*	75009	78951:	contig of 3843 bp in length
*	78952	78952:	gap of unknown length
*	78952	81039:	contig of 2088 bp in length
*	81040	81139:	gap of unknown length
*	81140	84921:	contig of 3782 bp in length
*	84922	85021:	gap of unknown length
*	85022	90311:	contig of 5290 bp in length
*	90312	90411:	gap of unknown length
*	90412	94521:	contig of 4114 bp in length
*	94526	94625:	gap of unknown length
*	94626	98229:	contig of 3604 bp in length
*	98230	98329:	gap of unknown length
*	98330	102008:	contig of 3679 bp in length
*	102009	102108:	gap of unknown length
*	102109	106715:	contig of 4607 bp in length
*	106716	106815:	gap of unknown length
*	106816	110900:	contig of 4085 bp in length
*	110901	111000:	gap of unknown length
*	111001	115834:	contig of 4834 bp in length
*	115835	115934:	gap of unknown length
*	115935	119585:	contig of 3651 bp in length
*	119586	119685:	gap of unknown length
*	119686	124375:	contig of 4690 bp in length
*	124376	124475:	gap of unknown length
*	124476	128981:	contig of 4506 bp in length

Query Match 11.6%; Score 36.6; DB 2; Length 209882;
 Best Local Similarity 52.3%; Pred. No. 1.7;
 Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY	4	CTGGGTATATCTTGAGCCCAATGAGCAACACCACTGGCC/TCTTCCTCTTGAGGCTA	63
Db	77673	CTCTGTTTGTCTCTGCTCTCATGTCCCACTCCAGATGGC/TTTATTACTTCTCTCTC	7773
QY	64	CACCCCTGCCCTTTCTTCCCTTGTAGTACTGTGCTGAGCTGATGAGCGCCACATCC	123
Db	77733	CCCCAGTAACACCTTTACACCGGATCTGTTCGATGGACAAATTTCTTAGGAACACCT	7779
QY	124	AAGCATGAGCCTGGACACATCTCCTCTGCCCCCTCC	158
Db	77793	TGGCCTGGCGCTGGCCAAAGATACCCCGCCGACCC	77827
RESULT 11			
RNO7628			
LOCUS	RNO7628	3688 bp	mRNA
DEFINITION	Rattus norvegicus mRNA for Elk channel 1.		linear
ACCESSION	AJ007628		
VERSION	AJ007628.1		
KEYWORDS	elk1 gene; potassium channel.		
SOURCE	Rattus norvegicus.		
ORGANISM	Rattus norvegicus.		
REFERENCE	1 (bases 1 to 3688)		
AUTHORS	Engeland, B.		
TITLE	Direct Submision		
JOURNAL	Submitted (03-JUL-1998) Engeland B., Zentrum fuer Molekulare Neurobiologie Hamburg, Institut fuer Neurale Signalverarbeitung, Martinistrasse, D-20246 Hamburg, GERMANY		
REFERENCE	2 (bases 1 to 3688)		
AUTHORS	Engeland, B., Neu, A., Ludwig, J., Roeper, J. and Pongs, O.		
TITLE	Identification of three rat potassium channel genes homologous to D. melanogaster elk		
JOURNAL	Unpublished		
FEATURES	location/Qualifiers		
source	1. 3688		
gene	/organism="Rattus norvegicus"		
CDS	/db_xref="taxon:10116"		
	/tissue-type="cortex"		
	99. 3152		
	/gene="elk1"		
	99. 3152		
	/gene="elk1"		
	/function="potassium channel"		
	/codon_start=-1		
	/evidence=experimental		
	/product="Elk channel 1"		
	/protein_id="CAA07587.1"		
	/db_xref="GI:3702616"		
	/db_xref="SPTREMBL:O89048"		
	/translation="MPYKGLAPONTFLDTIATRPDGTSHNPLANAQGRPGPIIV		
	CSGFCFLTGTGRTVEVMQKTCSCRFLETPTESEPALORLQKLEGHQHEHAIECFYRK		
	DGSAFCLLDMMPKIKNELGEVYLEFEEKIISOGGGLSPGIGHNNHNSIGRR		
	GASSRSTPRORNTYLHRLTGHGRGDDQSVANRVDPEKRPVEXKYASVGSGT		
	LILHYSLEKAWMDGLILATFYAAVTPPYVCRAGDDPTITSRHTLYSDIAEMETC		
	LDLILNRTTYSOSGOVVSAPNSIGLHATLATEVDLIALPDLILYVNITVLSLV		
	HLKTVLRLRLRLQLERY SOSAAVLLILMSVFLALAHMACWYVIGRRMEANL		
	DPLMDIGMLHELGKLEEPYVNSAGSPRSRAY IALFTTSLTSVGFANCAVNT		
	DAEFTISICTMILIGALMHAVFVGNVTAL IORMYRSLSYSRMDKLDFTLVHRLP		
	LKQMLFEPQTTMAVNSGIDANELLRPFLRLADIMHINRETIOLPLGASRGCL		
	RALSIIKTSFCAQEPFLRGLQALAHYVYSSGLEVDNLYATILTGKDLGADIT		
	PELQEPGAGACVYLKTSADYKALITGLOQLSRGLAEVRLYPEYVAARGLPND		
	LTLNLRQGSSENNNGIRFSRSPRLSQASDILIGSSDITLPSITETEGEMBPAGSKRP		
	RPLILPMSIPARPGSLVSLIGELIPFSALVSPSLSPSPALAGSSPSLHGPP		
	RGSAAMPOLITPLGTGPPGLDLPWIVDIGIEDSSNTAAEPTRFRCREPTTRSO		
	APLSGPTSLREIATEAEAEKVCYKRLNOEISRLNOVSEOLSRLELQVMDIARGLP		
	PSHPQDSTMLDLPDCHPDRPCISPHNSGPPGCLQNTLAVYHRAVSQVIEIGATPS		
	ELKSNMPPPPPEPDDPLGSPVPEASVLTLSLHNSQSGSDTFH"		
BASE COUNT	695 a	1200 c	979 g
ORIGIN			814 t

	Query Match	11.5%;	Score 36.2;	DB 10;	Length 3688;	
	Best Local Similarity	48.8%;	Pred. No. 2.5;	Mismatches 103;	Indels 0;	Gaps 0;
	Matches	98;	Conservative	0;	Mismatches	103;
OY	13 TCTGAGCCCATGCCATGTGCACACCACCTTGCCCTTCCTTACCTTAGGGCTACACCTCGCC	72				
Dd	3095 TCTGACCCCACAGCCTCCGTAAGCACAGCTTCCACTCTGGGTCACACAACTTCCACTGACC	3154				
OY	73 CTCTTTGGCCCCCTTTGATTACTGTGTGGCTGGAGTGAAGCGCACTCCCAAGCATGCC	132				
Dd	3155 CTGGCCCTTGGGCCACAGGCGCTGTCTGGGGGTGGCTTAATTACCTCCATCCAGGGAGAGC	3214				
OY	133 CTGGAACAATGTCCCTCCGCGCCGCCACACACACTGACAGGATTCAGTGGGGACTCGGGG	192				
Dd	3215 TGCGCTCTTGTGGCTCTTGGCTTGGGGGTGCACAGCTGCACACTGTGCTGGTGGCTCTGG	3274				
OY	193 GCAAAGGAGACTGTGTACAA 213					
Dd	3275 ATTCTGTGACTTTTTAACAA 3295					
RESULT 12						
LOCUS	ARI79196	3736 bp	DNA	linear	PAT 20-APR-2002	
DEFINITION	Sequence 10 from patent US 6326168.					
ACCESSION	ARI79196					
VERSION	ARI79196.1 GI:20220751					
KEYWORDS	.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
AUTHORS	1 (bases 1 to 3736)					
TITLE	Miyake,A., Mochizuki,S. and Yokoi,H.					
JOURNAL	Brain specific potassium channel protein					
FEATURES	Patent: US 6326168-A 10 04 -DEC-2001;					
	Location/Qualifiers					
	source	1..3736				
		/organism="unknown"				
BASE COUNT	706 a 1210 c 1000 g 820 t					
ORIGIN						
	Query Match	11.5%;	Score 36.2;	DB 6;	Length 3736;	
	Best Local Similarity	48.8%;	Pred. No. 2.5;	Mismatches 103;	Indels 0;	Gaps 0;
	Matches	98;	Conservative	0;	Mismatches	103;
OY	13 TCTGAGCCCATGCCATGTGCACACCACCTTGCCCTTCCTTAAAGGCTACACCTCGCC	72				
Dd	3132 TCTGACCCCAAGCCTCCGTAAGCACAGCTTCACAGTGGGGTCAACACAACTCCACTGACC	3191				
OY	73 CTCTTCTTGGCCCCCTTGTAGTACTGTGGCTGGAGTGAAGAGGCACTCCCAACATAGCC	132				
Dd	3192 CTGGCTTGGGCCACAGGCGCTGTCTGGGGTGGCTTAATTACTGTGCATCCAGGAGAGC	3251				
OY	133 CTGGAACAATGTCCCTCCGCGCCGCCACACACACTGACAGGATTCAGTGGGGACTCGGGG	192				
Dd	3252 TGCGCTCTTGGCGCTCTTGCGCTTGGGGGTGCACAGCACTGCACAGCTGGTGGTGGCTCTGG	3311				
OY	193 GCAAAGGAGACTGTGTACAA 213					
Dd	3312 ATTCTGTGACTTTTTAACAA 3332					
RESULT 13						
LOCUS	ABO22699	3736 bp	mRNA	linear	ROD 31-AUG-1999	
DEFINITION	Rattus norvegicus mRNA for BEC2, complete cds.					
ACCESSION	ABO22699					
VERSION	ABO22699.1 GI:5804789					
KEYWORDS	BEC2.					
SOURCE	Rattus norvegicus brain cdna to mRNA.					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					

REFERENCE 1 (sites)
 AUTHORS Miyake, A., Mochizuki, S., Yokoi, H., Kohda, M. and Furutachi, K.
 TITLE New ether-a-go-go K(+) channel family members localized in human telencephalon
 JOURNAL J. Biol. Chem. 274 (35), 25018-25025 (1999)
 MEDLINE 99386988
 REFERENCE 2 (bases 1 to 3736)
 AUTHORS Miyake, A., Mochizuki, S., Yokoi, H., Kohda, M. and Furutachi, K.
 TITLE Direct Submission
 JOURNAL Submitted (21-JUN-1999) Akira Miyake, Yamaguchi Pharmaceutical Co., Ltd., Molecular Medicine Laboratories, 21 Miyukigaoka, Tsukuba, Ibaraki 305-8585, Japan (E-mail: miyake@yamaguchi.co.jp, Tel: 81-298-52-5111 (ex. 3324), Fax: 81-298-52-5444)
 FEATURES
 SOURCE Location/Qualifiers
 1. 3736
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /tissue_type="brain"
 136..3189
 /codon_start=1
 /product="BEC2"
 /protein_id="BAAB3593.1"
 /db_xref="GI:5804790"
 /translation="MPVKGLLAPQNTFLDTIATREDGTHSNFLANAGPRGPIVY
 CSDGFCELTGYGRTEVMQKTSRFLGPEPTSEALQKALEGHEHAETCFYRK
 DGSAFWCLDMMPKINELGEVLFLEFKISOGSGGPGSGHGNHNNHNSIGRR
 GASRLRSTRONRTVLRHLGHRGDRDOSVKANSVPEKRPPEPKVAVSGSGRC
 LLHINSPKAVMDGILLATPEYAVTPYVPCFGRDDPTITSRTIVSDIANEMET
 LDITLNRRTYVSGQVVSAPRSIGVHTLATPFVFLIALPPDLIVFNITVTSLV
 HLKTVALLRLRLQLEKRYSCSAVALVLLMSVFLALHMAVCVYVIGRRMEAN
 DPLMDIGMLHEKRLIEPYVNGSAGPERSAYIALYFTLSLTVSGFNVICANT
 DAKRTISCTMLGALMAVFGVNTAIIORMSRSLYSRMDLQDFRVRHLRPP
 LKORLEYPOTAVNVSGIDANELLRDEPELRADIMHNRITLQPLRGASRGCL
 RALSINIKTSFCAFGELLRGDALAHYVCGSLVLELDNVTALIGRGDILGADIL
 PELGORGACAGVCIKTSADVKALTYGCLQOLSRCGLAEVLRKPEVVAFRAGLPD
 LITNLRGSENNIGRFSRSPRLSQASDYLSSSDTLPSITETEGMGSGSKPR
 RPLLPNLSAPRPGSLVSLGELPFSALVSPSLSPSPSPALAGSGSPSLHGP
 RGSANMKPQLLPPLTFCPPDLSPRIVIGIDEDSNTAJAPFRERKREPRTSRQ
 APLSGRSLRELTAELAEVEKVCRLNOEISRLNOSLSRLEKQVMDLQARLGP
 PSHPPDSTMLPDLCRPHQRPCISPHMSGPGLQNTLLAVHCPASVGVVEIGATPS
 ELRSSWPPPPSPDPPLGSPPEASPLTSPSLKHSQSSDTPFH"

BASE COUNT 706 a 1210 c 1000 g 820 t
 ORIGIN

Query Match 11.5%; Score 36.2; DB 10; Length 3736;
 Best Local Similarity 48.8%; Pred. No. 2.5;
 Matches 98; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 13 TCTGAGCCCATGCCATGTGCACACCACTGGCCCTCTTCTTAGGCGCTACACCCCTGCC 72
 DB 3132 TCTGAGCCCATGCCATGTGCACACCACTGGCCCTCTTCTTAGGCGCTACACCCCTGCC 3191
 QY 73 CTTTCTTTGGCCCTGTAGTTACTGTGCGGAGTGTGAGCGGCATCCCAAGCATGCC 132
 DB 3192 CTGGCCCTGGCCCAAGCGCTGTCTGGGGTGGCTTAATTAATTCCTGCGCATCCAGGAGGAGC 3251
 QY 133 CTGGAACAAATGCTCTCTCTCCCTCCACAGACAGTGCAGAGATTCACCTGGGAGCTCGGG 192
 DB 3252 TGGGCTCTCTGGGCTCTCTGGGCTGAGCAGACCTGCTGCTGGCTTGG 3311
 QY 193 GCAAGAGACAGCTGTGCACA 213
 DB 3312 ATTCTCTGAGACTTTTAAACA 3332

RESULT 14
 E32986 10732 bp DNA linear PAT 18-JUN-2001
 LOCUS E32986
 DEFINITION Gene encoding cellulose synthetizer.
 ACCESSION E32986
 VERSION E32986.1 GI:13022340

KEYWORDS JP 2000060568-A/1.
 SOURCE Vigna angularis.
 ORGANISM Vigna angularis
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustoids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.

REFERENCE 1 (bases 1 to 10732)
 AUTHORS Kotchi, M., Tomohiko, K., Shigeru, S. and Daisuke, S.
 TITLE Gene encoding cellulose synthetizer
 JOURNAL Patent: JP 2000060568-A 1 29-FEB-2000;
 COMMENT KOTCHI MIZUNO, MITSUI GIYOSAI SHOKUBUTSU BIO KENKYUSHO
 OS Vigna angularis
 PN JP 2000060568-A/1
 PD 29-FEB-2000
 PF 26-AUG-1998 JP 1998239998

PI KOTCHI MIZUNO, TOMOHIKO KATO, SHIGERU SATO, DAISUKE SHIBATA PC
 C12N15/09, A01H5/00, C07K16/40, C12N5/10, C12N9/10, C12P21/02// PC
 (C12N15/09, C12R1:91), (C12N5/10, C12R1:91), (C12P21/02, C12R1:91), PC
 C12N15/00,
 PC C12N5/00, (C12N15/00, C12R1:91), (C12N5/00, C12R1:91) CC
 FH Key Location/Qualifiers
 FT CDS (1)..(3375).

FEATURES
 SOURCE Location/Qualifiers
 1. 10732
 /organism="Vigna angularis"
 /db_xref="taxon:3914"
 BASE COUNT 3149 a 1212 c 2074 g 2046 t 2251 others
 ORIGIN

Query Match 11.5%; Score 36.2; DB 6; Length 10732;
 Best Local Similarity 26.9%; Pred. No. 2.4;
 Matches 39; Conservative 50; Mismatches 56; Indels 0; Gaps 0;

QY 20 CCAATGCGATGTGCACACCACTGGCCCTCTTCTTAGGCGCTACACCCCTGCTTCTT 79
 DB 9397 BYSSCTBYNSNYDAVSSRSCYCTCYTSCYRCAKCTBCNSTSAKRTTTTTT 9338

QY 80 TGCCCTCTTAGTACTGCTGAGTGTGAGTGTGAGCGGCACCTCCCAAGCATGGCTGGAAC 139
 DB 9337 YRCCTYSRCYTBCTCYTNCGRGCSRTTSNRVAYSCNCGNSDSTBCTTYDAC 9278

QY 140 AATGCTCTCTGCGCCCTCCACACGAC 164
 DB 9277 YTCYTVSDCCYTCNSTCYTCCSRK 9253

RESULT 15
 AC021632_1/c
 WPCOMMENT
 Sequence split into 5 fragments LOCUS AC021632 Accession AC021632
 Fragment Name Begin End
 AC021632_1 1 110000
 AC021632_2 100001 210000
 AC021632_3 200001 310000
 AC021632_4 300001 410000
 AC021632_5 400001 455450

Continuation (2 of 5) of AC021632 from base 100001 (AC021632 Mus musculus chromosome
 Query Match 11.5%; Score 36.2; DB 2; Length 110000;
 Best Local Similarity 50.9%; Pred. No. 2.3;
 Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 14 CTGAGCCCATGCGATGTGCACACCACTGGCCCTCTTCTTAGGCGCTACACCCCTGCC 73
 DB 33682 CTGAGCCCATGCGATGTGCACACCACTGGCCCTCTTCTTAGGCGCTACACCCCTGCC 33623

QY 74 TTTCTTTGCCCCCTTAGTACTGCTGAGTGTGAGTGTGAGCGGCACCTCCCAAGCATGGCC 133
 DB 33622 TGGCCTTTGGGCTTAGGCTCTCTGAGGCTAGGCTTTACACACTCTCCATCCAGGAGCGCT 33563

QY 134 TGGACAAATGCTCTCTGCGCCCTCCACACGACAGTGCAGAGATTCAGTGG 182

DB 33562 GGGCTCCTTGGCCCTTGGGCTCAGCAGCTGCCAGCTGACTGG 33514

Search completed: June 21, 2003, 03:52:44
Job time : 1086 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:31:01 ; Search time 80 Seconds

(without alignments)
270.437 Million cell updates/sec

Title: US-09-955-807-2

Sequence: 1 MLCSEPMPCAHPLGLFLG.....PHEPLTWGHLINHYSSASH 105

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671380

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

SPTREMBL.21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	15.3	678	5 Q960P9	Q960P9 drosophila
2	89	15.3	743	5 Q9VBM6	Q9VBM6 drosophila
3	79.5	13.7	684	4 Q9H6M8	Q9H6M8 homo sapien
4	79.5	13.7	684	4 Q96BD1	Q96BD1 homo sapien
5	79.5	13.7	826	10 Q9FSK3	Q9FSK3 oryza sativ
6	79.5	13.6	261	10 Q9AUN2	Q9AUN2 oryza sativ
7	76.5	13.1	206	10 Q8RUD7	Q8RUD7 oryza sativ
8	76.5	13.1	1660	5 Q9KNO	Q9KNO leishmania
9	75.5	13.0	549	13 Q9W639	Q9W639 xenopus lae
10	74	12.7	220	11 Q9CVP2	Q9CVP2 mus musculu
11	73.5	12.6	132	2 Q9RPO0	Q9RPO0 chlamydia t
12	73.5	12.6	357	4 Q9NZ14	Q9NZ14 homo sapien
13	73	12.5	135	9 Q9MC36	Q9MC36 bacterioph
14	72.5	12.5	364	12 Q9WRN7	Q9WRN7 macaca mula
15	72.5	12.5	554	10 Q23691	Q23691 arabidopsis
16	72	12.4	1089	16 Q9X308	Q9X308 streptomyc

17	71	12.2	110	2 Q53356	Q53356 streptomyc
18	71	12.2	494	10 Q8S6A8	Q8S6A8 oryza sativ
19	71	12.2	1217	4 Q9ULU5	Q9ULU5 homo sapien
20	70	12.0	497	16 Q93JB9	Q93JB9 streptomyc
21	70	12.0	679	10 Q9FY53	Q9FY53 arabidopsis
22	70	12.0	1620	16 P96285	P96285 mycobacteri
23	69.5	11.9	141	4 Q8REG2	Q8REG2 homo sapien
24	69.5	11.9	147	4 Q9H5U7	Q9H5U7 homo sapien
25	69.5	11.9	434	16 Q9RVL8	Q9RVL8 deinococcus
26	69.5	11.9	860	16 Q9PE21	Q9PE21 xylella fas
27	69.5	11.9	883	2 Q55102	Q55102 streptomyc
28	69	11.9	287	5 Q20754	Q20754 caenorhabdi
29	69	11.9	328	16 Q9A7U1	Q9A7U1 caulobacter
30	69	11.9	365	5 Q8T4A4	Q8T4A4 drosophila
31	69	11.9	447	5 Q9VUJ3	Q9VUJ3 drosophila
32	69	11.9	507	13 Q9W7K3	Q9W7K3 brachydanio
33	69	11.9	588	10 Q9SDC0	Q9SDC0 oryza sativ
34	68.5	11.8	194	5 P91787	P91787 drosophila
35	68.5	11.8	202	16 Q55623	Q55623 synechocyst
36	68.5	11.8	323	11 Q9ECW6	Q9ECW6 mus musculu
37	68.5	11.8	335	2 Q9F8U5	Q9F8U5 streptomyc
38	68.5	11.8	399	16 Q8YK50	Q8YK50 anabaena sp
39	68.5	11.8	1235	5 Q9YFE3	Q9YFE3 drosophila
40	68	11.7	586	5 Q9W1Y2	Q9W1Y2 drosophila
41	68	11.7	1400	16 P96419	P96419 mycobacteri
42	67.5	11.6	220	4 Q96S11	Q96S11 homo sapien
43	67.5	11.6	277	4 Q96ID9	Q96ID9 homo sapien
44	67.5	11.6	333	4 Q96NT7	Q96NT7 homo sapien
45	67.5	11.6	352	4 Q9BYE7	Q9BYE7 homo sapien

ALIGNMENTS

RESULT 1

ID	Q960P9	PRELIMINARY:	PRT:	678 AA.
AC	Q960P9:			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DR	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DE	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
GN	LD40883P.			
OS	CG11849.			
OC	Drosophila melanogaster (Fruit fly).			
CC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OX	Ephydroidea; Drosophilidae; Drosophila.			
RN	NCBI_TaxID=7227;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RA	STRAIN=BERKELEY;			
RA	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,			
RA	Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,			
RA	Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,			
RA	Nunoo J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,			
RA	Yu C., Lewis S.E., Rubin G.M., Ceiniker S.			
RL	Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: AY051924; AAK93348.1; -			
DR	FLYBase: FBgn0039286; CG11849.			
SQ	SEQUENCE 678 AA; 71914 MW; 821E4FC920048BE2 CRC64;			

Query Match 15.3%; Score 89; DB 5; Length 678;
Best Local Similarity 32.4%; Pred. No. 0.4;
Matches 22; Conservative 15; Mismatches 23; Indels 8; Gaps 3;

QY	18 LGLHPAL--SLPLVTVAGVMSATPKH--GLECCPPAPPAVTGFTDSGAKETVSODK 73
DB	330 MGLNLPPIPTPIPKVTSPIRSTRPMQSLAQRPPIPSAPLPIRSPSSSLD----EK 385
QY	74 RSQGHFWC 81
DB	386 NAATYMC 393

```

RESULT 2
Q9VBM6 PRELIMINARY: PRT: 743 AA.
AC 09VBM6:
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DE CG11849 protein.
OS CG11849.
OC Drosophila melanogaster (Fruit fly)
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe C.M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Adair J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson R., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evansgela C.C., Ferreira C., Fertala S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ideyama C.,
RA Jalali M., Kalish K., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Malsbina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Stryker R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2193(2000).
DR EMBL: AE003751; AAF56411.1;
DR FlyBase: FBgn0039286; CG11849.
SQ SEQUENCE 743 AA; 7863 MW; B32C5BDD4999B8F6 CRC64;

Query Match 15.3%; Score 89; DB 5; Length 743;
Best Local Similarity 32.4%; Pred. No. 0.44;
Matches 22; Conservative 15; Mismatches 23; Indels 8; Gaps 3;

OY 18 LGGLHAL--SLPLVTVAGVMSATPKH--GLECCPAPAPAYATGFGSGAETYSQDK 73
DB 330 MGLNLPYPTTIPMGVTSPPINSSTPOHMSQLAQTPTIPAPILPSTPSTPGSLD---EK 385
OY 74 RSQGHWTG 81
DB 386 NAAWYMW 393

```

```

RESULT 3
Q9H6M8 PRELIMINARY: PRT: 684 AA.
AC 09H6M8:
DT 01-MAR-2001 (TRENBLREL. 16, Created)
DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
DE CDNA: FLJ22061 fis, clone HEP10105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobayake N., Inagaki Y., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Odayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isonaga T., Sugano S.,
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR EMBL: AK025714; BAB15227.1;
DR InterPro: IPR000759; Adrxn_reductase.
DR InterPro: IPR000866; ER_target.
DR InterPro: IPR001327; FAD_pyr_redox.
DR PRINTS: PR00419; ADXRDTASE.
DR PRINTS: PR00368; FADPNR.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
KW FAD; Flavoprotein; Oxidoreductase.
SQ SEQUENCE 684 AA; 7761 MW; 99134E1F6299CAE CRC64;

Query Match 13.7%; Score 79.5; DB 4; Length 684;
Best Local Similarity 29.7%; Pred. No. 4;
Matches 35; Conservative 13; Mismatches 53; Indels 17; Gaps 5;

OY 2 LGYSEPPCAHPGLFL-LGLHPALSLP-----LVYTVAGVMSATPKHGLSQC-----P 49
DB 1 MGSAAAPLWGPPLLLAIALHPALSVPRRDYCVLAGAGLQMAVFLQAGRDYAVFE 60
OY 50 PAPPAPAYATGFGSGAETYSQDKRSQGHWTCTALPMPILTWG-----LRNHSVA 103
DB 61 RAPRPG-SPTTRPRHKKLSINKRYTKNAERNLHMDNLSLSHPRLFRHSYA 117

RESULT 4
Q96BD1 PRELIMINARY: PRT: 684 AA.
AC 096BD1:
DT 01-DEC-2001 (TRENBLREL. 19, Created)
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
DE 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
DE Hypothetical (77.7 kDa protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC015726; AAH15726.1;
DR InterPro: IPR000886; ER_target.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 684 AA; 77730 MW; 8109EE9565A0D676 CRC64;

Query Match 13.7%; Score 79.5; DB 4; Length 684;
Best Local Similarity 29.7%; Pred. No. 4;
Matches 35; Conservative 13; Mismatches 53; Indels 17; Gaps 5;

OY 2 LGYSEPPCAHPGLFL-LGLHPALSLP-----LVYTVAGVMSATPKHGLSQC-----P 49

```

Db 1 MGSLAALWGPGLIALIALHPALSVPPRRDYCVLAGAPLQMAVFTQGRGRDYAVE 60
 QY 50 PAPPVATGTGDSGAKETVSODKRSQGHWTCTLALPHPLTWVGH-----LRNHVSSA 103
 Db 61 RAPRRG-SFTFRYPRHKLISINKRTYGRANAEPENLRHDMNSLSHDPRLFRHYSRA 117

RESULT 5

Q9FSK3 PRELIMINARY: PRT: 826 AA.
 AC Q9FSK3;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Hypothetical 92.0 kDa protein..
 GN H0806H05.17.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OC NCBI_TaxID=4530;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Han B., Feng Q., Huang Y.C., Chen Z.H., Zhou B., Li Y., Zhu J.J.,
 RA Tang Y.S., Zhao Q., Liu Y.L., Mu J., Yu Z., Fan D.L., Chen L.,
 RA Wang Q.D., Zhang L., Lu Y.Q., Yu S.L., Zhu J., Liu X.H., Hu X.,
 RA Lei H.Y., Zhang Y.J., Wang R., Li C., Lu Y., Chen X.C., Zhang Y.,
 RA Hu H., Jia P.X., Li T., Qian Y.M., Ying K., Hong G.F.,
 RA "Oryza sativa indica (Guangjia14) genomic DNA, chromosome 4, BAC
 RT clone: H0806H05."
 RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AL442113; CAC09484.1;
 DR Hypothetical protein.
 KW SEQUENCE 826 AA; 92046 MW; 12295155B874E6A8 CRC64;

Query Match 13.7%; Score 79.5; DB 10; Length 826;
 Best Local Similarity 25.4%; Pred. No. 4.9;
 Matches 25; Conservative 12; Mismatches 36; Indels 37; Gaps 6;

QY 7 PMPAHPLGLFLGLHPALSLPLVTVVAGVMSATPKH-----GLEQC-----PPAPPA 55
 Db 14 PLPLRP-----IHQASILEAAVASSILLADPEVVAAPACSSGSPPLPPPPQA 66
 QY 56 VTGFTGSGAKETVSODKRSQGH-----WCTL-----ALPHPLTW 92
 Db 67 SP--TMDSTAASPSDKQARWPTMAGDGRWMDIIMHTLLGDAVGNLMSM 118

RESULT 6

Q9AUN2 PRELIMINARY: PRT: 261 AA.

AC Q9AUN2;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical 28.5 kDa protein..
 GN OSJNBA0058E19.19.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OC NCBI_TaxID=4530;
 RN [1]
 RP SPOUNCE FROM N.A.
 RA Spiegel L.A., King L., Kirchoff K.A., de la Bastide M., Preston R.R.,
 RA Nascimbeno L.U., Vill M.D., Baker J.P., Miller B., Cunnius D.M.,
 RA Kait K.H., Rodriguez S., Santos L., Zutavern T., Ballia V.S.,
 RA Shah R.S., Bahret A., Bal H.P., O'Shaughnessy A., Dedhia N.N.,
 RA McCombie W.R.;
 RA "Genomic Sequence for Oryza sativa, Nipponbare Strain, Chromosome X,
 RT clone OSJNBA0058E19, complete sequence."
 RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC083943; AK13144.1;

DR InterPro: IPR005213; HGMP.
 DR Pfam: PF03578; HGMP; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 261 AA; 28496 MW; DEC112EC16E4B05B CRC64;

Query Match 13.6%; Score 79; DB 10; Length 261;
 Best Local Similarity 26.5%; Pred. No. 1.7;
 Matches 36; Conservative 9; Mismatches 43; Indels 48; Gaps 5;

QY 12 HPLGLFLGLHPALSLPLVTVVAGVMS--ATPRHGLEQCPAPPPVATGTGDS----- 63
 Db 4 HNAAGVILLGAGSCLPVPGPVAVGVSVLLSMARHGWKLCHVPRVTVVMTSDQSITTS 63
 QY 64 -----GAKETVSODKRSQ-----HTWC-----TLALPHPLTWV-- 93
 Db 64 NSRALGVGAEOCTQEMIIQNTTGLPSPRLHSMVLPPLGVIITFAGLPSPLTGVFA 123
 QY 94 -----GHLRNH 99
 Db 124 YMAIGTAVDGHRLRL 139

RESULT 7

Q8RUV7 PRELIMINARY: PRT: 206 AA.

AC Q8RUV7;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical 22.5 kDa protein..
 GN OSJNBA0091J06.11 OR OJ1341F06.4.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OC NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=CV. NIPPONBARE;
 RA McCombie W.R., de la Bastide M., Spiegel L., Preston R., Ferraro K.,
 RA Kuit K., Nascimbeno L., Zutavern T., Ballia V., Bell M., Baker J.,
 RA Santos L., Miller B., Katzenberger F., Muller S., King L., Yang C.,
 RA Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;
 RA "Genomic sequence for Oryza sativa, Nipponbare strain, clone
 RT OSJNBA0091J06, from chromosome 10, complete sequence."
 RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RL [12]

RP SEQUENCE FROM N.A.
 RA STRAIN=CV. NIPPONBARE;
 RA McCombie W.R., Spiegel L., de la Bastide M., Preston R., Ferraro K.,
 RA Kuit K., Nascimbeno L., Zutavern T., Ballia V., Bell M., Baker J.,
 RA Santos L., Miller B., Katzenberger F., Muller S., King L., Yang C.,
 RA Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;
 RA "Genomic sequence for Oryza sativa, Nipponbare strain, clone
 RT OJ1341F06, from chromosome 10, complete sequence."
 RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC113338; AAM08655.1;
 DR EMBL: AC116926; AAM08884.1;
 KW Hypothetical protein.
 SQ SEQUENCE 206 AA; 22473 MW; 7906E311D793FB3C CRC64;

Query Match 13.1%; Score 76.5; DB 10; Length 206;
 Best Local Similarity 31.0%; Pred. No. 2.4;
 Matches 26; Conservative 7; Mismatches 30; Indels 21; Gaps 4;

QY 11 AHPGLFLGLHPALSLPLVTVVAG--VMSATPRHGLEQC-----PPAPPA 54
 Db 14 AHAHVHT-LPETHAPKPSPIAFVTGDLTYMEASPEPLPCWAHRCFALIHCLPPAPP 72
 QY 55 AVTGFTGDSGAKETVSODKRSQGH 78
 Db 73 SRTG-----TGAPSORRQARRIAAH 92

RESULT 8
 Q9NKN0 PRELIMINARY: PRT: 1660 AA.
 AC Q9NKN0: 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE L5204.7.
 GN L5204.7.
 OS Leishmania major.
 OC Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RA Myler P.J., Sisk E., Hixson G., Kiser P., Rickel E., Hasselbrock M.,
 RA Cawthra J., Margolin F., Sunkin S., Stuart K.D.,
 RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC005941; AF34294.1;
 DR InterPro: IPR002965; P_itch_extensn.
 DR PRINTS: PRO1217; PRICHEXTENS.
 DR PROSITE: PS00227; TUBULIN; UNKNOWN.1.
 SQ SEQUENCE 1660 AA; 170277 MW; 38CD35AE74346116 CRC64;

Query Match 13.1%; Score 76.5; DB 5; Length 1660;
 Best Local Similarity 34.1%; Pred. No. 21;
 Matches 31; Conservative 7; Mismatches 36; Indels 17; Gaps 5;
 Oy 6 EPMPCAHPLGLFLGLHPALSLPLVYVAGVMSATPKHGLECCP-----PAPPPAVT 57
 Db 360 EPRP-ASPTWPL---LAAITPLASTALAVTSABPRSLGALPLSSASRSPSPPAV- 414
 Oy 58 GFTGDSGAKETVYSDKRSOGHTWCTIALP 88
 Db 415 --PSSSSAPTLA--FASPTPHRCQAPPPP 441

RESULT 9
 Q9W639 PRELIMINARY: PRT: 549 AA.
 AC Q9W639: 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Xsmad4.
 GN XSMAD4.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=99223549; Pubmed=10207044;
 RA Masuyama N., Hanafusa H., Kusakabe M., Shibuya H., Nishida E.,
 RT "Identification of two smad4 proteins in Xenopus: chier common and
 RT distinct properties."
 RL J. Biol. Chem. 274:12163-12170(1999).
 DR EMBL: AB022721; BAB7514.1;
 DR HSSP: Q13485; 1DD1.
 DR TRANSFAC: T04153;
 DR InterPro: IPR001132; Dwarf1n.
 DR InterPro: IPR003619; Dwarf1n_A.
 DR InterPro: IPR004863; MH1.
 DR Pfam: PF03165; MH1.1.
 DR Pfam: PF03166; MH2.1.
 DR SMART: SM00523; DWA; 1.
 DR SMART: SM00524; DWA; 1.
 SQ SEQUENCE 549 AA; 59855 MW; 72431FD6AC674150 CRC64;

Query Match 13.0%; Score 75.5; DB 13; Length 549;

Best Local Similarity 26.7%; Pred. No. 8.4;
 Matches 27; Conservative 8; Mismatches 37; Indels 29; Gaps 4;
 Oy 23 PALSLPLVYVAGVMSATPKHGLECCPAP-----PA-----VTGTGDSGAK 66
 Db 219 PSTSQP-----ASLAAHNDGLSLINAVPFGQGNFSAQAPATVHNHNSITWTGSRTA 273
 Oy 67 ETVSDKRSOGHTWCTIALPPLTWG-----LRNH 99
 Db 274 YTPNMSHHNGLQHHPMPHGHVYVHNEIAFQPPISNH 314

RESULT 10
 Q9CVP2 PRELIMINARY: PRT: 220 AA.
 AC Q9CVP2: 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 1700110N18R1k protein (Fragment).
 GN 1700110N18R1k.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; Pubmed=11217851;
 RA Kawai J., Shingawa A., Shidata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., Ring B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Resole G., Quackenbush J.,
 RA Schirli L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya K., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski B., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK007165; BAB24879.1;
 DR MGD: MGI:1920819; 1700110N18R1k.
 FT NON-TER
 SQ SEQUENCE 220 AA; 24067 MW; 9D8780481470406A CRC64;

Query Match 12.7%; Score 74; DB 11; Length 220;
 Best Local Similarity 27.2%; Pred. No. 4.6;
 Matches 28; Conservative 12; Mismatches 47; Indels 16; Gaps 4;
 Oy 7 PMPANHLGLFLGLHP--ALSLPLVYVAGVMSATPKHGLECCPAPPPAVTGTGDSG 64
 Db 56 PABC-----LGVNPDFOVTPHCTFTTADPNMSRPGSLGTLQTPAPPPAS-----ESW 104
 Oy 65 AKETVSODKRSOGH---TWCTIALPPLTWGHLRNHVSAS 104
 Db 105 HYPASQVSPSYSHMDMTLRNHNPRANVHNHNNHNTAGS 147

RESULT 11
 Q9RPO0 PRELIMINARY: PRT: 132 AA.
 AC Q9RPO0: 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

[2]
 RN SEQUENCE FROM N.A.
 RP SPECIES=Macaca mulatta rhadinovirus 26-95;
 RC STRAIN=MACACA MULATTA RHADINOVIRUS ISOLATE 26-95;
 RX MEDLINE=20173730; PubMed=10708456;
 RA Alexander L., Denekamp L., Knapp A., Auerbach M.R., Damania B.,
 Desrosiers R.C.;
 RT "The primary sequence of rhesus monkey rhadinovirus isolate 26-95;
 RT sequence similarities to Kaposi's sarcoma-associated herpesvirus and
 RT rhesus monkey rhadinovirus isolate 17577.";
 RL J. Virol. 74:3388-3396(2000).
 DR EMBL; AF083501; AAD21391.1; -
 DR EMBL; AF210726; AAF60043.1; -
 DR InterPro; IPR001346; IRF.
 DR Pfam; PF00605; IRF; 1.
 DR SMART; SM00348; IRF; 1.
 SQ SEQUENCE 364 AA; 41189 MW; D8A48195C39DD018 CRC64;

Query Match 12.5%; Score 72.5; DB 12; Length 364;
 Best Local Similarity 35.1%; Pred. No. 11;
 Matches 20; Conservative 5; Mismatches 25; Indels 7; Gaps 2;

QY 32 TVAGVMSATFKHGEGCPAPPAVGTGDSGAKETVSQDKRSQGTWCTIALPH 88
 DB 125 TVLNTLEACVHGLEGCTPLPPAPA--EADGAARSVYARAR---LATVAPPH 174

RESULT 15

023691 PRELIMINARY; PRT; 554 AA.
 AC 023691;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Hypothetical 60.9 kDa protein.
 GN T19D16.24.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
 RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
 RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome 1 BAC T19D16 genomic sequence.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U95973; AAB54931.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 554 AA; 60902 MW; 8188B41P9AE27BBA CRC64;

Query Match 12.5%; Score 72.5; DB 10; Length 554;
 Best Local Similarity 32.4%; Pred. No. 17;
 Matches 22; Conservative 3; Mismatches 30; Indels 13; Gaps 1;

QY 21 LHPALSLPLVTVAGVMSATPKHGLEQCP-----APPAYVTGTGDSGAK 67
 DB 232 LPPPPPLPMVAVRKGVAAAPLPPTGTALPPPPPLPMMAKGVAAPPPPGARGGLGAKK 291
 QY 68 TVSQDKRS 75
 DB 292 VTSKLRKS 299

Search completed: June 17, 2003, 10:36:55
 Job time : 84 secs